

DOCUMENT RESUME

ED 066 975

FL 003 451

AUTHOR	Toma, Peter P.; And Others
TITLE	Optimization of Systran System.
INSTITUTION	LATSEC, Inc., La Jolla, Calif.
SPONS AGENCY	Rome Air Development Center, Griffiss AFB, N.Y.
REPORT NO	RADC-TR-72-73
PUB DATE	Apr 72
NOTE	195p.
EDRS PRICE	MF-\$0.65 HC-\$6.58
DESCRIPTORS	*Computer Programs; Data Processing; Dictionaries; *English; *Linguistic Theory; *Machine Translation; *Russian; Semantics; Syntax

ABSTRACT

This report describes an optimization phase of the SYSTRAN (System Translation) machine translation technique. The most distinctive characteristic of SYSTRAN is the absence of pre-editing; the program reads tapes containing raw and unedited Russian texts, carries out dictionary and table lookups, performs all syntactic analysis procedures, and prints out translated texts without human direction. The report begins with a schematic introduction to SYSTRAN, followed by an extensive discussion of improvements effected during optimization in the dictionary, parsing system, and semantics. Improvements in other areas, i.e., the translation of homographs, adverbs, proper names, numerals, adnominal genitives, verbs plus objects, sequence of tenses, abbreviations, and articles, are also reported. A discussion of programing specifics in improving the efficiency of SYSTRAN processing is included, followed by remarks on SYSTRAN and the controversy concerning machine translation in general. Appendixes provide further details on the operation of the program. (Author/VM)

U. S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE
PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION
POSITION OR POLICY

RADC-TR-72-73
Final Technical Report
April 1972



OPTIMIZATION OF SYSTRAN SYSTEM

LATSEC, Incorporated

Approved for public release;
distribution unlimited.

Rome Air Development Center
Air Force Systems Command
Griffiss Air Force Base, New York

ED 066975

FL003 451

OPTIMIZATION OF SYSTRAN SYSTEM

Dr. Peter P. Toma
Dr. Ludek A. Kozlik
Mr. Donald G. Perwin

LATSEC, Incorporated

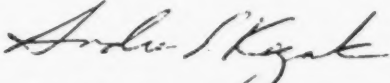
Approved for public release;
distribution unlimited.

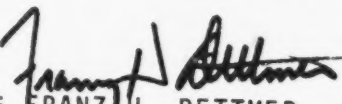
FOREWORD

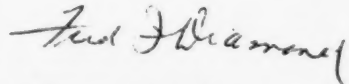
This Final Technical Report has been submitted by LATSEC, Incorporated, 7852 Ivanhoe Avenue, La Jolla, California, under Contract F30602-71-C-0091, Job Order No. 30080000, Task 04, for Rome Air Development Center, Griffiss Air Force Base, New York. Report covers the period 5 October 1970 to 5 October 1971. Other contributors to the report are the following: Alexandra Andrich, Diana Aruffo, Judith Garrett, Dr. Paul Krasnopolksy, James Peabbles, Kathleen Roberts, Larissa Toma and Chuck Walrad. The RADC project engineer was Andrew Kozak (IRDT).

This technical report has been reviewed by the Office of Information (OI) and is releasable to the National Technical Information Services (NTIS).

This technical report has been reviewed and is approved.


Approved: ANDREW S. KOZAK
Technical Evaluator


Approved: FRANZ H. DETTMER
Colonel, USAF
Chief, Intel & Recon Division

FOR THE COMMANDER: 

FRED I. DIAMOND
Actg. Chief, Plans Office

ABSTRACT

This report describes an optimization phase of the operational machine translation system designed by LATSEC, INC., and designated SYSTRAN. The most distinctive characteristic of SYSTRAN is the absence of pre-editing. Whenever SYSTRAN's operation is initiated on the computer, it reads in tapes containing raw and unedited Russian texts, carries out dictionary and table lookups, performs all syntactic analysis procedures, and prints out translated texts without human direction. Optimization during this phase was primarily focussed, as directed in the Work Statement, on update and correction. The former task include further development of homograph routines, improved handling of adnominal genitives, numbers, proper names, abbreviations, sequences of tenses, and adverbs and sophistication of article insertion, in addition to the larger task of implementing the Segment Analyzer Program. The latter task, dictionary optimization, was directed at correcting human errors, using special dictionary print programs to pinpoint errors in existing entries, and at updating the dictionary, using diagnostic messages provided by the computer during update to prohibit introduction of new errors. These messages indicate specific reasons for rejecting invalid transactions. Optimization in the area of semantics was limited, since the major emphasis of the Work Statement were dictionary and parsing improvement. The optimization effected improves the handling of polysemantic lexical items by examining their context. This

report, however, reflects SYSTRAN's readiness to incorporate greater semantic advances and describes directions which semantic optimization must necessarily take in the future.

TECHNICAL EVALUATION

Final Technical Report, "Optimization of SYSTRAN System" on Contract F30602-71-C-0091, LATSEC, Inc.

The optimization effort described in subject report consisted of the following tasks in order of descending emphasis: (1) verification and modification for inclusion in the SYSTRAN dictionary system of (a) an additional 50,000 compound S&T dictionary entries directly absorbed from the Mark II Russian Master Dictionary, and (b) an additional 70,000 S&T stem entries; (2) output translation improvements based upon analysis of homographs, adverbs, adnominal genitives, numerals, proper names, sequences of tenses, abbreviations and article insertion; (3) further sentence-wide improvements on the syntactic level (parsing) and semantic level (contextual solutions); (4) implementation of means for improving efficiency of SYSTRAN processing on IBM 360/65 equipment.

The above tasks have been carried out concurrently with maintenance of a fully operational Russian-English MT capability throughout the reporting period.

Dealt with effectively under Task (1) is a critical and demanding requirement of MT -- compilation of a large MT lexical and grammatical data base for extensive text coverage. The following are represented in SYSTRAN dictionaries: physics, electronics, computer hardware and data processing, aviation and space, mechanical engineering, biology and medicine, metallurgy, military sciences, chemistry, earth sciences and mathematics. The total data base includes approximately 144,000 stem entries, 160,000 compound entries and 3,300 idiomatic and phrasal constructions. Subject report explicates the use of a SYSTRAN dictionary update subsystem, but not reflected are the often overlooked consistent efforts by those linguists who have proofread each of the added 130,000 entries, providing equivalents and syntactic coding as required for MT operations. Task (2) is addressed to solutions of severe target-language problems caused by the above listed elements of the source language. Detailed classifications, especially for treatment of homographs, together with large sample lists of class and subclass members (including illustrative computer translation), reflect essential linguistic components for an operational MT system. Task (3) deals with automatic sentence-structuring of larger components, as clauses and phrases, required both for definition of logical relationships and word order rearrangement; a Segment Analyzer Program has been implemented for these purposes. Contextual resolution of polysemantic words has proceeded on a modest scale for priority reasons. Generally, resolution is achieved by taking the intersection of assigned semantic properties and syntactic relationships of a set of lexical items in text, where at least one item is polysemantic in the target language (see p. 40-42). A modicum of contextual information is utilized for treatment of pronouns during

translation (see p. 118, 119). For Task (4), measures to improve SYSTRAN processing efficiency have been implemented in the area of automatic dictionary data processing (see p. 92-94). The current SYSTRAN System output rates are reported in the range of 250-300,000 English words per hour of elapsed time.

The System operates on an IBM 360/65 with a minimum of 512K bytes of main storage. Peripherals are: two IBM 2314 direct access facilities; two IBM 2401 tape drives (7-channel, 800 bpi magnetic tape); IBM 2540 card reader for SYSIN; IBM 1403 printer for SYSOUT: 360 MVT or MFT OS. The system consists of 27 operational programs and 19 utilities (total of 51,373 instructions), and is programmed in 360 Direct Assembler Language.



ANDREW S. KOZAK
Technical Evaluator

TABLE OF CONTENTS

	<u>Page</u>
I. INTRODUCTION	1
1. GENERAL	1
2. A SCHEMATIC INTRODUCTION TO SYSTRAN	2
a. An Overview	3
b. SYSTRAN Supporting Systems	6
c. SYSTRAN Translation Processing	8
II. OPTIMIZATION	18
1. INTRODUCTION	18
2. DICTIONARY OPTIMIZATION	19
3. PARSING OPTIMIZATION	26
a. Segment Analyzer Program	27
b. Rearrangement	31
c. A Comparison of Human and SYSTRAN Parsing	32
4. SEMANTICS	40
III. FURTHER IMPROVEMENTS EFFECTED DURING OPTIMI- ZATION	43
1. INTRODUCTION	43
a. Improving the Translation of Homographs	43
b. Adverbs	55
c. Proper Names	63
d. Numerals	65
e. Adnominal Genitives	67

TABLE OF CONTENTS

	<u>Page</u>
f. Verbs + Objects	69
g. Sequence of Tenses	70
h. Abbreviations	72
i. Articles	80
IV. PROGRAMMING SPECIFICS IN IMPROVING THE EFFICIENCY OF SYSTRAN PROCESSING	92
1. SELECTION OF L. S. COMPOUNDS WITHIN TOPICAL GLOSSARIES	92
2. LEXICAL ROUTINES: IMPROVED HANDLING OF MULTIPLE-MEANING WORDS	94
V. SYSTRAN AND THE CONTROVERSY ABOUT THE FEASIBILITY OF MACHINE TRANSLATION	95
VI. SYSTRAN AND THE CONTROVERSY ABOUT THE STATE OF THE ART	102
VII. SYSTRAN'S POSITION	112
VIII. THE FUTURE OF MACHINE TRANSLATION	116
BIBLIOGRAPHY	121
APPENDIX I	
"EXPERIMENTS IN INTERNATIONAL TELECON- FERENCING": ORIGINAL TEXT	
APPENDIX II	
RUSSIAN TRANSLATION	
APPENDIX III	
ANALYSIS OF "EXPERIMENTS IN INTERNATIONAL TELECONFERENCING"	

TABLE OF ILLUSTRATIONS

	<u>Page</u>
THE SYSTRAN TRANSLATION PROCESS	5
360-65 SYSTRAN TRANSLATION PROCESSING	14
SYSTRAN SUPPORTING SYSTEMS	15
CREATION OF L. S. COMPOUND DICTIONARY	16
GRAPH ANALYSIS OF HUMAN PARSING (Figure 1)	34
GRAPH ANALYSIS OF SYSTRAN PARSING (Figure 2)	35
SEGMENT ANALYSIS, HUMAN (Figure 3)	38
SEGMENT ANALYSIS, SYSTRAN (Figure 4)	39

SECTION I

INTRODUCTION

1. GENERAL

This report describes an optimization phase of the SYSTRAN (System Translation) machine translation technique. During this phase some important milestones were reached in the development of an operational machine translation system, which may have an impact on current linguistic theories. Included is a description of the basic tasks performed during this period. In addition to its translation capability, the developed system is also capable of gathering data for its own further improvement and of testing grammars.

This report takes into consideration the fact that the goals, methods, and feasibility of machine translation have been controversial issues during the past ten years. SYSTRAN is presented as a typical example of an open-ended system which develops a computer intermediate language before carrying out any translation. SYSTRAN's parsing technique is compared with that of humans. In addition, the approaches taken at semantic and contextual levels are illustrated.

Before the report enters into any of the above mentioned considerations, illustrations accompanied by explanations are used to present SYSTRAN in toto. It is hoped that, by this means, a better understanding of the System and its optimization will be afforded. Chapters I through V are devoted to the

description of the work carried out under the contract. Each contains translated samples showing how the tasks envisaged within the work statement of this optimization phase have been fulfilled.

It is significant that this refinement effort was directed primarily at the systematic correction of errors. These errors had been made mostly at the level of human dictionary coding. They have been detected by analyzing large amounts of computer output. In order to detect and rectify these errors, it was necessary that the computer utilize its existing codes in translating texts. Only the computer itself can produce all the logical consequences of any particular coding schema. No human being, no matter how well versed in Russian grammar, can correctly assess the consequences of the coding--or lack of coding--he specifies. It is possible, however, with training, to achieve a very high degree of error recognition in translated output and to deduce, using one's own knowledge of the parsing rules, what sorts of codes, by their absence or improper assignment, resulted in the translation error.

2. A SCHEMATIC INTRODUCTION TO SYSTRAN

This section is designed to give a basic explanation of how the SYSTRAN translation process operates. Three different schemata are included, giving (1) an overview of the entire process from input of the original Russian text to final machine-translated output as it is received by the post editor, (2) a closer look at the actual machine translation process within the

IBM 360/65, and (3) an examination of the special dictionaries formulated and used by SYSTRAN.

a. AN OVERVIEW. Any text selected for translation is typed on an MT/ST equipped with both Cyrillic and Roman characters, as well as numerals. (The SYSTRAN System can also receive input directly from punched cards. This particular input procedure is not illustrated here.) The MT/ST operator, while familiar with the Cyrillic alphabet, need have no knowledge of the Russian language per se. No editing whatsoever is required, so the operator simply types whatever appears in the text to be translated, indicating at appropriate points how much space should be left for the later inclusion of graphic material (e.g., diagrams and charts) which cannot be typed. This material can then be inserted in the translated output.

Each MT/ST cartridge can contain up to 24,000 letters, or approximately 3,000 Russian words. As the text is typed on the MT/ST equipment, an MT/ST cartridge automatically receives a particular bit configuration for each letter typed. It is this information that is converted from the cartridges to 360 tapes. The conversion consists of the automatic transformation of the bit configuration on the cartridges into 360 bit configurations. A 360 tape can hold the contents of approximately 600 MT/ST cartridges, but any number of cartridges (or even half of a cartridge) can be immediately converted and translated.

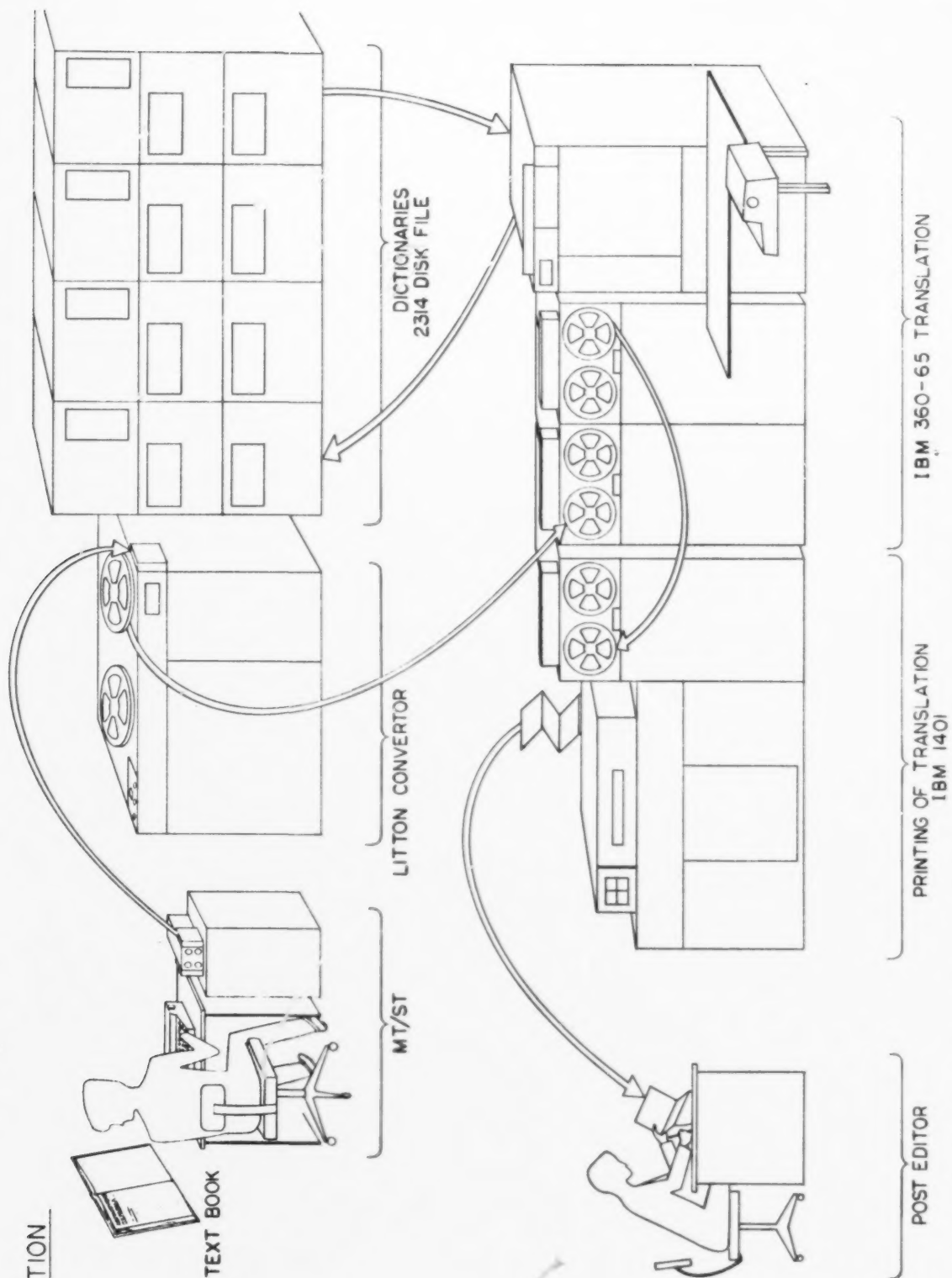
The converted 360 tapes serve as direct input to the translation system. The LOADTXT program reads the text into the

computer while the 360 processes the text, the main frame taking full advantage of the 2314 disk file system attached to it. The disk file arrangement is specially suited to the translation process. It is used by the high speed core primarily for dictionary and compound expression lookups. The information contained in these dictionaries, glossaries, etc., is constantly utilized throughout the translation process.

The disks also contain the entire translation system prior to its activation. Each section (program, routine, etc.) is called by the control program, INITCALL, and loaded into the high speed core as it is needed. This whole process of sequential control is performed entirely automatically without human interference. In other words, as soon as the text is loaded in and INITCALL is activated, SYSTRAN carries out any and all tasks necessary for dictionary lookups, structural passes, etc., until a complete syntactic analysis has been achieved and the total input has been translated.

The translated product, still on 360 tape, is then written into a magnetic tape in a 1401 compatible format and subsequently printed out by an IBM 1403. The printed output contains upper and lower case letters and suitable punctuation and spacing. The post editor therefore receives a translation on which he can perform tasks similar to those of any editor working on the product of a human translator.

TRANSLATION



b. SYSTRAN SUPPORTING SYSTEMS. The most important part of machine translation, in that it is the most basic part, is comprised of its dictionaries. SYSTRAN's are stored in the 2314 disk file. The disk dictionaries necessary to SYSTRAN translation processing are created by utilizing a series of supporting systems. These systems can be described as follows:

- (1) Stem Dictionary Update system maintains the master tape of stem dictionary entries.
- (2) Idiom/Limited Semantic (L. S.) Dictionary Update system maintains the master tape of Idiom and L. S. dictionary entries.
- (3) Creation of L. S. Compound Dictionary system creates the L. S. compound dictionary on disk and updates the stem master dictionary with this information.

The operational procedure of these systems is schematized in Chart II, parts 1 through 3, and is explained in the following paragraphs.

Stem Dictionary Update. (See Chart II, part 1.) Updating transactions for the dictionary are keypunched and read first by an editing program (DEDIT). All codes entered on the card are edited for correctness and formatted for computer processing. The edited transactions are then sorted to the master dictionary sequence (stem and stem number) and are ready to update the master dictionary. The update takes place in two operations. In the first pass of the master tape, all stem entries except

additional forms of infinitive verbs are updated by the master update program (DUPDT). The infinitive verb carries a table of the additional forms, and when the infinitive is updated, all dictionary codes for the infinitive are written out for each additional form. These records are then sorted in master dictionary sequence and the intermediate master dictionary created by DUPDT is updated with these entries by the additional verb stem update program (AVSPASS). A listing of the new master stem dictionary is prepared by a dictionary print program (DPRINT).

Idiom/L. S. Compound Dictionary. (See Chart II, part 2.)

The initial load of the master dictionaries on disk do not contain L. S. compound information but are used in creating the compound disk dictionary. The first two programs (LOADTXT and MDL) of the SYSTRAN translation system are used to retrieve dictionary information from the disk for association with the individual words of each L. S. expression. These dictionary codes are analyzed by a program (CLS3) to create an individual L. S. compound record. (All L. S. programs are identified by the letters CLS plus a number.) The principal word of the expression is assigned by this program and a supplementary word updating record is written to a separate file. The individual compound records are sorted to principal word sequence and read by the next L. S. compound creation program (CLS4). This program creates a master compound record combining all L. S. compounds which have the same principal word and topical glossary into a single compound record. As these compound records are

written on disk, their address is included in an index record and written on disk for each principal word indicating the location of the compounds for the different topical glossaries for this word. An updating record of this index is written in a separate file. The supplementary and principal word index updating records are read by another stem dictionary updating program (DUPDTLS) and the address of the index record and principal and/or supplementary word indicators are attached to the stem dictionary master. Again, as in updating the stem dictionary by DUPDT, the L. S. information for the additional forms of the verb are written to a file and sorted to dictionary sequence. The sorted records are read and the final new master stem dictionary is created by the AVSPASS program. A repeat of the Disk Dictionary Load system is performed to establish the translation disk dictionaries with L. S. compound information. All disk dictionaries are now ready for SYSTRAN translation processing.

c. SYSTRAN TRANSLATION PROCESSING. While there are inherent difficulties associated with substituting a computer for a human translator, primarily due to the computer's inability to "understand," the computer's limitations in this respect are overcome by SYSTRAN in utilizing the machine's great speed, its unlimited capacity for storing source material, and its ability to perform repetitious and tedious tasks without error. The logic of the SYSTRAN translation process can be divided into three primary phases: table lookup, syntactic analysis in the

source language and synthesis into the target language. (See Chart III.)

The tables carry information necessary for the accurate translation of any source text. Many of them are accessed through the use of addresses to variable lengths areas which carry the expanded information on disks. The codes and addresses are affixed to the source text and carried from this phase into syntactic analysis and synthesis.

LOADTXT executes the instructions for input and work during this phase. As each word enters the high speed core, it is matched against a table of high frequency words which also contains the first word of all idioms. (See Chart III, High Frequency and Idiom Lookup). Locating these words and separating from low frequency words saves considerable search time, since the HF/Idiom Table is kept in the high speed core and therefore does not have to be brought in from auxiliary disk storage. Moreover, the meaning and grammar codes of high frequency words are stored separately in a table accessed by offset or address linkage. The addresses are stored with each word in the HF/Idiom Table and are attached to source text words when matches are found. These text words are then stored in a high frequency file.

Source text words which are not matched in the above process are sent directly to a low frequency file for the main dictionary lookup (MDL) program. MDL calls in the words from the low frequency file and accumulates them on disk. The

portions of text words, along with portions of stem tables (which are also on disks) are read into memory. (See Chart III, MDL.) Stems are stored on three different disks, according to length (1-3 characters, 4-7 characters, 8-24 characters). Each entry, a source language stem word, has attached to it paradigmatic set codes (PST) indicating the endings permitted for that stem. There is also an address at each stem pointing to the disk location of additional grammar and meaning information and to Limited Semantic (L. S.) information if the entry may function in an L. S. expression.

After all the words in the low frequency file have been processed by MDL and supplied with appropriate grammar and meaning codes, they are sorted and put into a merged file with the words from the high frequency file. At this point, all the words with their codes are put back into their original sequence in the sentence and the sentence is ready for syntactic analysis.

The programs which perform the syntactic analysis are initiated and sequentially ordered by INITCALL. INITCALL first calls GETSENTN to initiate the analysis of the words as they function in the particular sentence held in the analysis area. (See Chart III, Sentence Analysis Area.) The sentences are called individually from the merged file into the analysis area in memory. The first 88 bits (11 bytes) of each word's 160 bytes contain constant information which will not be modified by the structural passes.

Words which have been determined by LOADTXT as components

of L. S. expressions are marked as having been translated. The principal word of the expression will carry the target meaning of the entire expression.

The structural passes (STRPASS0 - STRPASS4) indicate the results of their activity in the 1192 bits remaining for each word in the sentence analysis area. The information placed in these bytes is not static and may be re-evaluated during any succeeding pass.

STRPASS0 detects homographs and resolves their part of speech ambiguities by using special subroutines to analyze the relationships of homographs to other parts of the sentence. For example, it may decide that a noun/verb homograph functions in a particular sentence as an infinitive because it is immediately preceded by a word which has an infinitive requirement. STRPASS0 also marks definite objects by setting appropriate bits at the source text words.

As STRPASS1 moves from right to left (sentence-end to sentence-beginning), each part of speech (POS) code encountered initiates subroutines to set switches indicating that word's position and potential function.

STRPASS2 determines the function of punctuation marks. Commas, semi-colons, parentheses, quotation marks, dashes, etc., all carry POS codes, just as individual words do. These codes indicate functions of punctuation marks, and yield additional information contributing to marking structures not revealed in STRPASS1.

STRPASS3 locates subjects and predicates. It uses two primary tools: POS codes and a hierarchical search criteria based on the probability of certain parts of speech becoming subjects or predicates more commonly than others. The predicate is determined first; then the subject search begins. Subjects and predicates must agree in number and gender (whenever applicable).

STRPASS3 also analyzes commas with indeterminate function. Since the location of subject and predicate often aids in the discernment of embedded clauses, some commas may now be recognized as subsentence unit (SSU) separators. In addition, some commas are clearly seen to be enumerative, in the case of multiple subjects and/or predicates.

The subject-predicate search in STRPASS3 may also occasion re-evaluation of a noun marked by STRPASS1 as an object. If no other subject candidate is found, that object will be re-evaluated as the subject if a finite verb has been found during the predicate search.

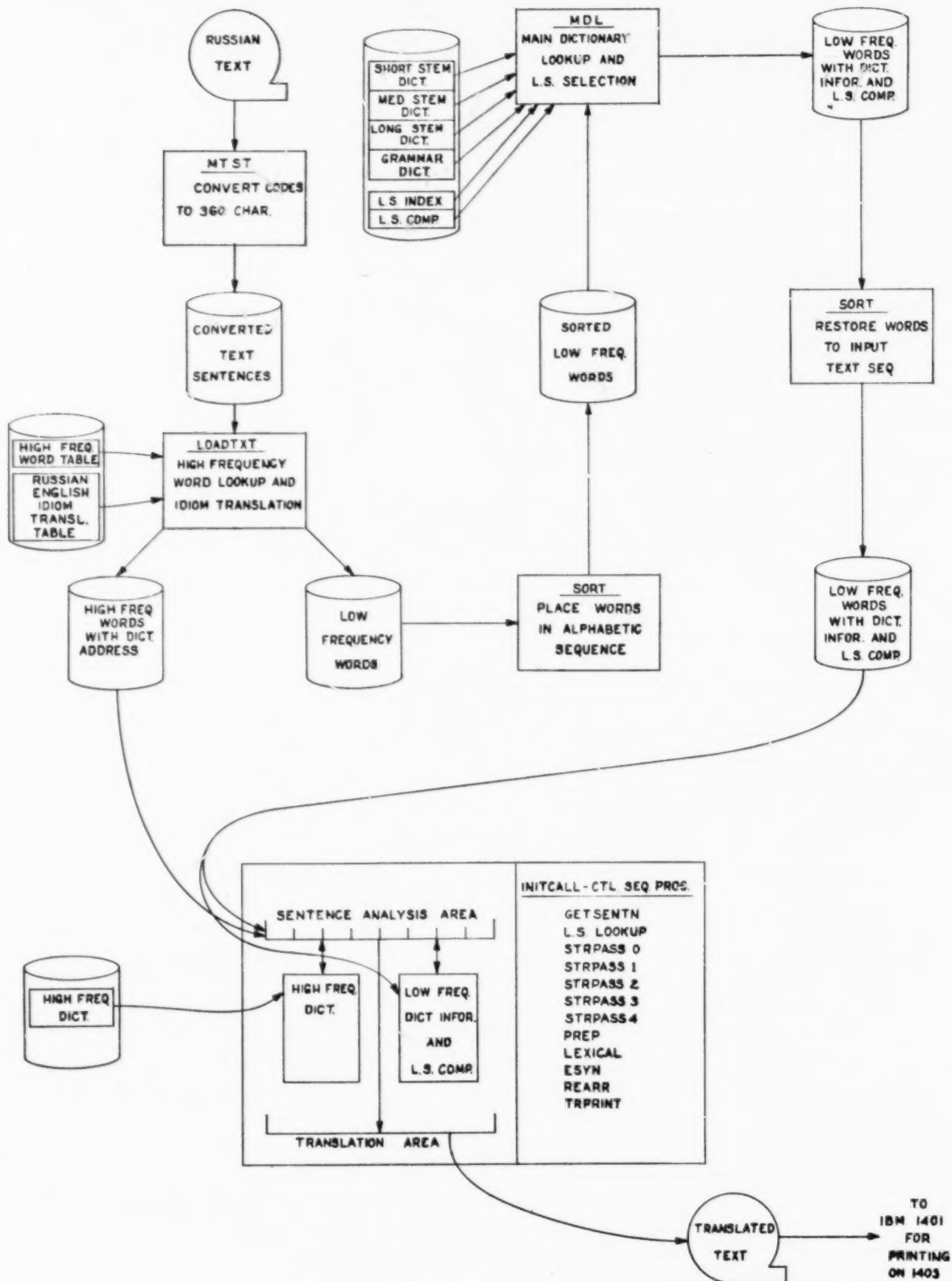
STRPASS4 looks at the entire sentence for words which have not been previously recognized. STRPASS4 continues to operate until it has marked all words as having governing or subordinate functions. When words both govern and are governed in turn, that situation will also be indicated.

The third and last major part of the translation process is the synthesis into the target language. This is accomplished, basically, by branching on each part of speech. For example, the routines which translate verbs into English use tables in

memory. Selection of these tables depends in part on whether the verb is active or reflexive or whether or not the auxiliary TO BE is used in any conjugation. Each table looks for tense, aspect, person, number, conditionality, negation, etc., to determine which English auxiliary or pronoun should be used with any particular verb.

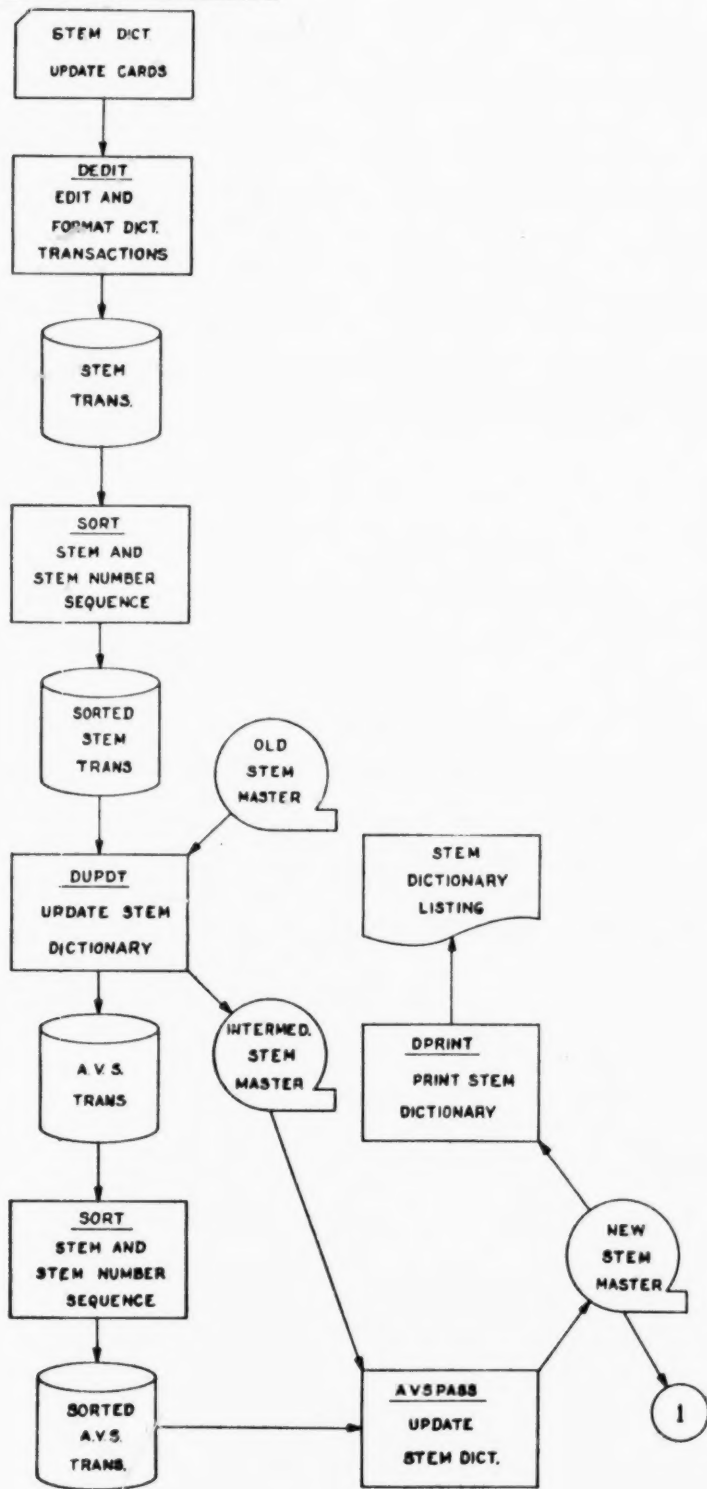
There are also special routines for inserting English articles (Russian has no articles) and for ensuring the proper translation of prepositions. The rearrangement program takes into consideration the unique word order requirements of the target language. It contains rules governing word order in English which, with the grammar codes, enable the program to establish correct sequences of words and to direct that this specified word order be printed in the translated output.

360-65 SYSTRAN TRANSLATION PROCESSING

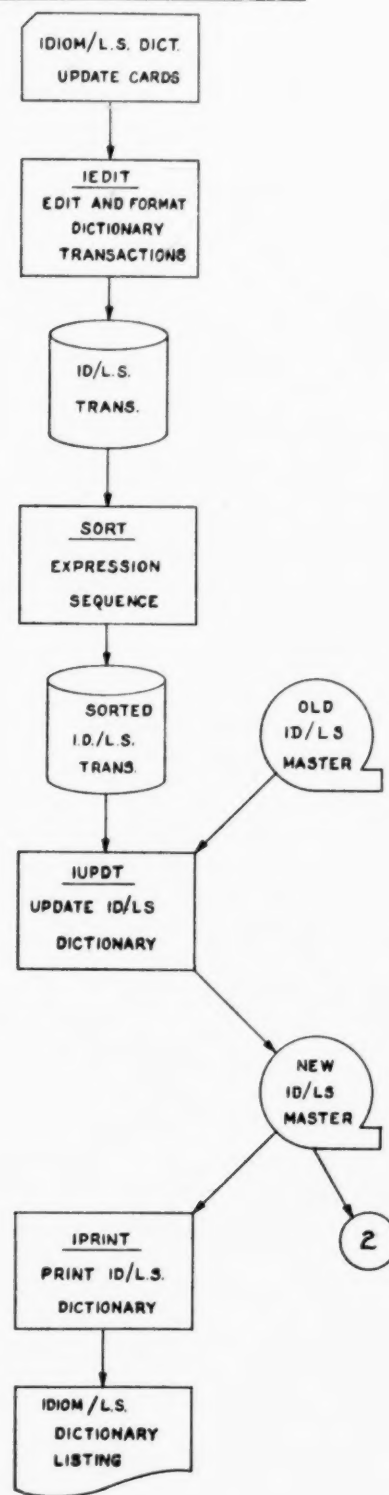


SYSTRAN SUPPORTING SYSTEMS

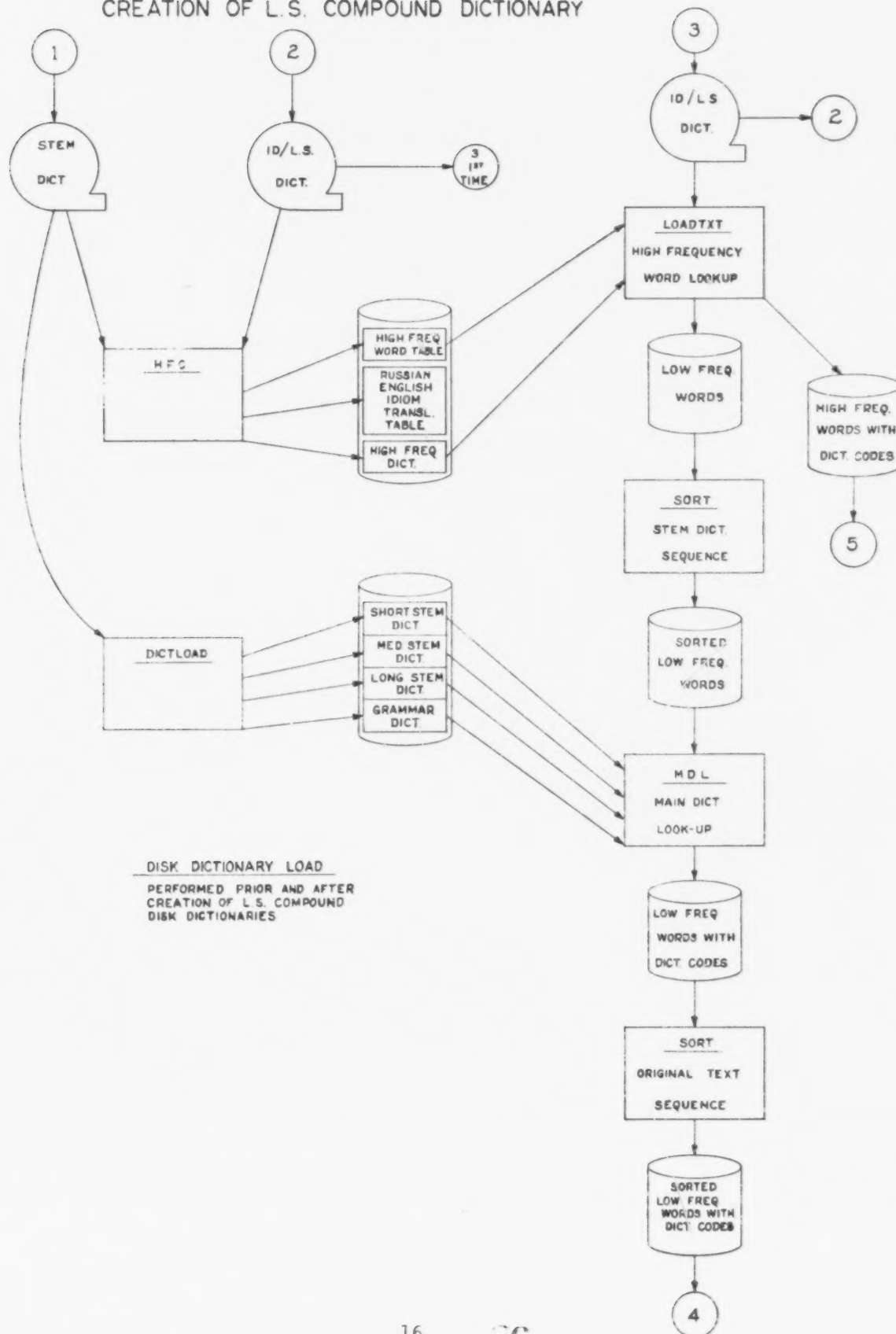
STEM DICTIONARY UPDATE

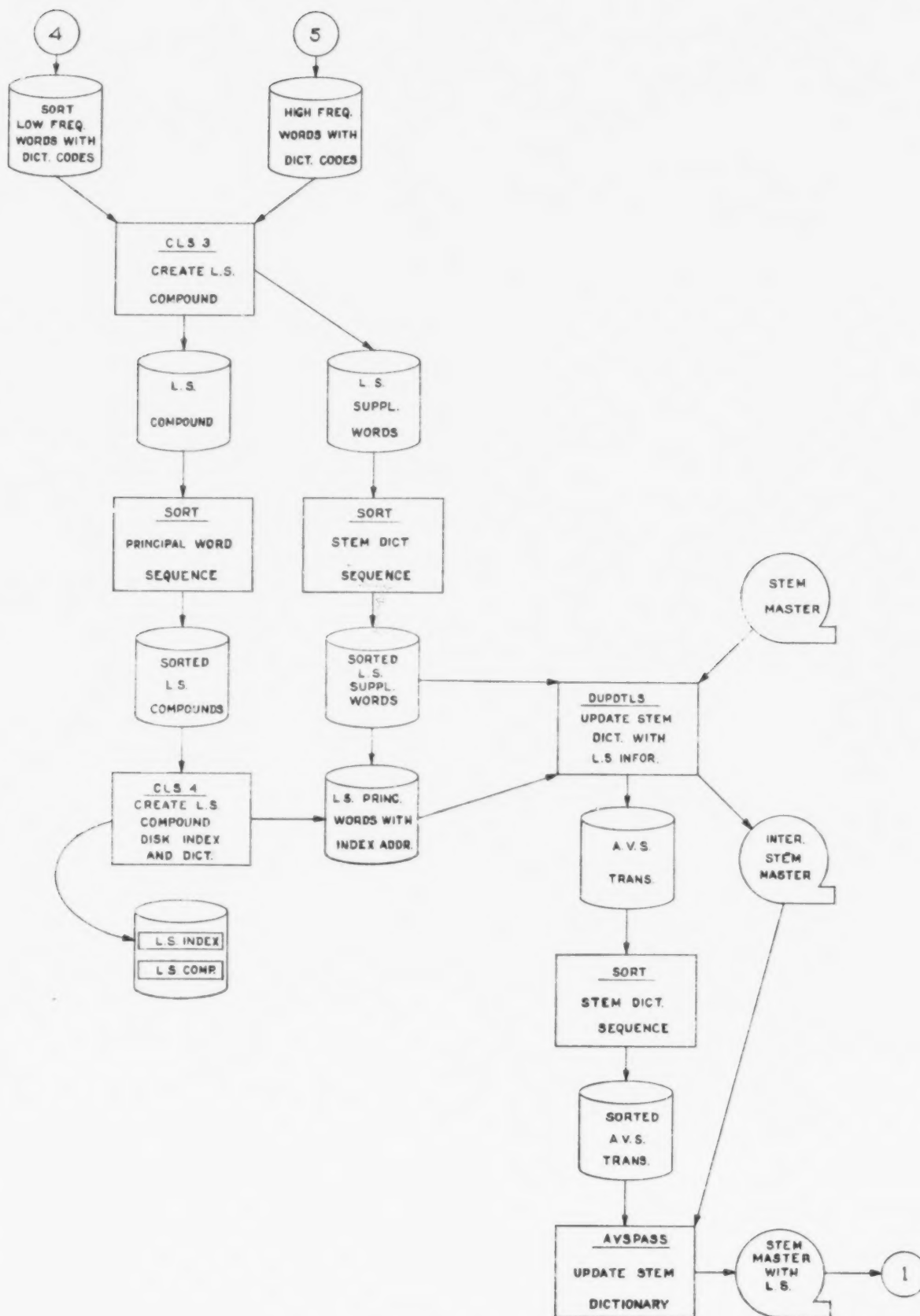


IDIOM/L.S. DICTIONARY UPDATE



CREATION OF L.S. COMPOUND DICTIONARY





SECTION II

OPTIMIZATION

1. INTRODUCTION

Optimization of SYSTRAN dictionaries during this phase was primarily directed at the correction of errors introduced at the human level during coding. The key aspect of dictionary optimization has been SYSTRAN's ability to check possible errors and print diagnostic messages regarding the types of errors it encounters.

Improvements in SYSTRAN's parsing abilities were dependent to a large extent on correction of dictionary codes, since these codes refer not only to meaning information, but also to important grammatical information, such as parts of speech and government requirements of entries. In addition, the SEGMENT ANALYZER Program was instituted to analyze clauses and phrases by syntactic levels.

The Work Statement de-emphasized semantics in favor of lexical and syntactic improvements. Therefore, the work effort in this area was directed primarily at handling polysemantic lexical items. Nonetheless, it is felt that a great deal of further investigation into the area of semantics is called for at this time. SYSTRAN is now ready to implement more sophisticated semantic criteria in its recognition grammar. A survey of the types of improvements envisaged in this area is presented in section eight of this report ("The Future of Machine Translation").

2. DICTIONARY OPTIMIZATION

SYSTRAN dictionaries have been considerably improved. Grammar codes and synthesis codes were systematically checked and whenever necessary updated or corrected. Meanings for about 70,000 stem entries and 50,000 compound entries were checked. During the systematic checking and correction, full advantage was taken of built-in SYSTRAN dictionary update features. The SYSTRAN dictionary update program, which accepts corrections or new entries, checks on possible errors on two levels and prints a diagnostic message whenever it may be assumed that any coding or keypunch error has occurred.

The first level of checking is performed directly on the keypunched transactions by an edit program (DEDIT and IEDIT) which immediately eliminates from the update process those records containing detected errors. The second level of checking (DUPDT) occurs during the actual update of the dictionary master tape and ensures the proper modification of the master file. Prior to replacing the dictionaries on disk that are used during translation, these diagnostic messages are analyzed, corresponding correction transactions are prepared, and the dictionary master tape is again updated. This update subsystem of SYSTRAN is a powerful feature which assures that the working dictionaries of this operational system can only be improved at each update, without running the risk of introducing new errors.

In the following, a list of messages will be given which the SYSTRAN dictionary updating subsystem may use any time an

unacceptable update is presented to it. These messages contain many abbreviations which refer to the SYSTRAN coding manual and coding techniques. It would be outside the scope of this report to elaborate on all these messages and their significance as far as update is concerned. The intent here is to give a general illustration of the means by which an update subsystem can prevent the incorporation of inaccurate data in the SYSTRAN dictionaries.

SYSTRAN DIAGNOSTIC MESSAGES

A. DEDIT (used in stem dictionary update)

<u>MSG</u>	<u>REASON</u>
BAD TRAN CODE	TRC(col. 28) not A,D,R, or M.
BAD CARD CODE	CARD NUM (col. 27) Not 0 thru 9 or A thru F.
BAD POS-BPQ	POS-BPQ not in valid program table.
HF MARKED FOR/NOT FF	DSN (col. 25) not 0.
HF STEM L. GT 14	Stem (col. 1-24) cannot be longer than 14 char. for HF word.
BAD PST CODE	PST (col. 42-44) may not be blank for TRC = A or R.
	PST, if present for noun/adj., must be 001 thru 255.
NANP CODE ERROR	Must be blank, - or 1-3.
ANIM CODE ERROR	Must be blank, - or 1-5.
NO PREVIOUS 1R TRAN	For TRC = 2R, there must be a previous 1R card.
PNAD CODE ERROR	Must be blank, - or 1-2.
BAD TG CODE	Must be 0-9 or A-D.
NO ENGLISH MEAN.	Meaning must be present for TRC=A/R.
SEQ CANNOT BE BLANK	Seq. must be present for TRC=M/R.
ILLEGAL SEQ	Seq. - 1st pos. must be 0-9, 2nd pos. must be blank or 0-9.
BAD DPQ	DPQ is not 0, 1, 2, or 3.
BAD SYN CODE	If DPQ=0, SYN must be blank or 000.
	=1/2, 1st pos. must be blank or 0-9. 2-3 pos. must be in valid table. =3, must be in valid table.

<u>MSG</u>	<u>REASON</u>
BAD ART CODE	Must be blank, - or 0-7.
BAD V/ADCODE	If DPQ = 0/1, must be blank, - or 0. =2, must be blank, - or 0-5. =3, must be blank, - or 0-9.
BAD VB CODE	If DPQ ≠ 3, must be blank, - or 0. =3, must be blank, - or 0-3.
NO FOL REF MN-EOF	Meaning card specifies REF, but is last card in input.
NO FOL REF MN-ERR	Meaning card specifies REF, but next card is not for same stem and/or TG.
FOL CD - PROCESSED	The card following preceding error card is printed for reference - it is not rejected unless noted in following error message.
NOT PROC - PREV ERR	Meaning card specifies REF, but following ref. meaning card was rejected, so this card is also rejected.
TRAN CODE ERROR	TRC must be A or R for card A-F or D for cards B-F.
NO PST COL 42-44	PST col. 42-44 may not be blank for card A.
NO AV STEM	Cards B-F have blank AV Stem.
PREP CODE ERROR	Prep. code must be blank, 99, - or 00-68.
SYN/VAD GT NUM WDS	1st pos. SYN for noun/adj or VAD for verb is greater than num. words in meaning.
INVALID STEM NUM	DSN must be 1st Pos 0-3, 2nd Pos 0-8.
BAD DSN FOR POS	Dictionary Stem Number and POS do not correspond.
POS 8X, BAD DSN	POS 8X is inappropriate with this DSN.
NO GENDER	Gender must be supplied for this to be translated.
ILLEGAL HOM CODE	Homograph code must be numerical, not alphabetic.
NO FF PST	PST codes must be supplied for full forms.
INVALID VBLS CODE	No S, P, or O.
NO CONT. LONG STEM	Stem longer than 24 characters with no following card.
BAD LONG STEM DSN3X	The verb is not allowed because it has more than 24 characters.
TRC = M WITH REF IND	Any change in nonreflexive verb affects reflexive form.

Additional messages may be specified by the coder if the message isn't already available on the coding sheet. The following

samples of messages SYSTRAN has used are self-explanatory:

NO ADD, DEL, REP

BAD SYMBOL

UNDEF SYMBOL

NO MULT FMT

BAD MEAN SYMBOL

NO SECOND CARD

B. IEDIT (used in Idiom/L.S. dictionary update)

MSG

REASON

NO R CARD
ERR CD CODE
REJ-ENG CARD ERROR

No Russian card for this English card.
Card code not A thru D.
Russian card rejected because of previous English card error.

NO E CD FOLLOWING
UNMATCH SEQ

No English card for this Russian card.
English card does not match on card seq. num. with Russian card.

BLANK RUS. EXP.
1st WORD GT 14

Russian expression not entered.
Length of 1st word in Russian idiom longer than 14 char.

RUS. EXP. TOO LONG

After separating commas with spaces, the Russian expression is greater than 55 char.

OVER 8 WORDS IN EXP

Russian L.S. expression contains more than 8 words.

BAD PO CODE
BAD PDA CODE

L.S. PO must be blank or 0-7.
L.S. PDA must be blank or 0-2.

C. DUPDT (used in Stem dictionary update)

REJ-

REASON

MSTR EXISTS
MSTR IS AVS

TRC=A for existing stem entry.
Transaction for add verb. stem. - may only be updated by "Verb Endings Coding Sheet."

NO MASTER STEM

Stem does not exist for this transaction.

NO OA TRAN

There was no previous transaction card OA to establish stem entry.

CHG TO NEW REC

Change to a stem just being entered.

REJ-REASON

DUP TRANS	Multiple transaction card for updating same section of stem
UNMATCH SEQ	Meaning seq. does not exist for this change.
MATCH ON SEQ	Meaning seq. exists for this add.
REFL MEANING	Cannot delete reflexive meaning. Reflexive meaning is deleted automatically when deleting non-reflexive meaning.
AV STEM NF	Deletion for non-existent add. verb stem reference.
AV STEM EXISTS	Add. verb stem reference already exists for add transaction.
REFL VS NONR	TRC interferes with reflexive-non-reflexive relationship.
AVS STEM	Don't delete stem if next card is an AVS card for that stem.
NO TARGET MEANING	Meaning must be present for TRC=A/R/M.

LIST OF ABBREVIATIONS USED IN MESSAGES

TRC	= Transaction Code
DSN	= Dictionary Stem Number
A	= Add
R(Rep)	= Replace
M	= Modify
DPQ	= Primary Part of Speech in Target Language
PST	= Paradigmatic Set Table Codes
SYN	= Synthesis
VAD	= Verb-Adverb
AV	= Additional Verb
REF	= Reflexive
EOF	= End of File
MN	= Meaning
GT	= Greater Than
NANP	= Noun-Adjective/Noun-Participle
FF	= Full Form
PNAD	= Participle Modifying Neither Verb Nor Adjective
PO	= Prepositional Object
PDA	= Prepositional Object Definite Article

In addition to updating the dictionary, a good deal of correction of existing entries was necessary to ensure that all entries carried all the proper grammar codes. The size of the dictionary made it requisite to discover automatic means to pinpoint sections of the dictionary requiring intensive review, to appropriately code deficient entries, and then to update the dictionary with the corrected codes. For this purpose, Special Dictionary Print Programs were designed and implemented to be used when certain entries were discovered which lacked necessary codes, when new codes or coding methods are devised to be incorporated in certain categories of entries, or when consistent erroneous coding has been discovered and can be isolated for editing by a few preliminary tests (i.e., RMD nouns ending in AM, AMI, AX).

The following special dictionary prints were programmed, run, and coded in these categories:

ENTRIES LACKING VITAL CODES

<u>DPRT</u>	<u>PURPOSE</u>
01	SYSTRAN verbs with a zero PST.
23	Print record with no target meaning.
52, 56	SYSTRAN DSN 1X, PST 000.
57	SYSTRAN verbs with PST E000000000 incorporated from Library of Congress records.
63	SYSTRAN verbs with syntax area either blank or all zeros.
70	Gender and number equal to FF.

NEW CODES OR CODING METHODS

<u>DPRT</u>	<u>PURPOSE</u>
04	Stems with Homograph indicators.
09	POS 14 (Proper nouns, Common names).
14	All Russian hyphenated stems.
17	All POS 18.
18	All BPQ 53.
27-45	Noun PST's converted to Full-form; PST 198 for possible conversion to the PST 204.
47	PST 198.
51, 67	Slashed English meanings to ensure that both a slashed and a non-slashed entry exist.
53	All PST's converted to full forms.
54	POS 24.
63	BPQ 30, to assign BPQ 3D, 3E, or 3F on the basis of its tendency to be used as a predicate.

ERRONEOUS CODING DISCOVERED

<u>DPRT</u>	<u>PURPOSE</u>
13	RMD nouns ending in AM, AMI, AX.
15	Stems ending in NEIW.
62	Full forms with FF in the PST.
71	Nouns with genitive requirement.

3. PARSING OPTIMIZATION

By utilizing 160 bytes for each word in any sentence under analysis, SYSTRAN allows access to all grammatical and semantic information pertinent to that word in any possible context.

The information contained in the 160 bytes attached to each word of the sentence (as briefly discussed later in this chapter) expresses the syntactic interconnections of the sentence in a manner understandable to the computer. The algorithms of the various programs which supply the codes held in these bytes are guided by the principles of the SYSTRAN grammar. This grammar grew out of the necessity of applying the computer to natural languages.

As Dykema points out, grammar began as a philosophical inquiry into the nature of language.¹ The most methodological and global inquiry into any system can be carried out by the computer if a suitable software system exists. SYSTRAN has all the prerequisites of such a system. Its system is modular, has several subsystems composed of major programs subdivided by minor programs, and further subdivided by major and minor routines. These programs and routines may, whenever necessary, call generalized or specialized subroutines, which in turn use nested routines for special inquiries and special coding tasks.

The SYSTRAN grammar breaks away from many traditional grammatical concepts. For example, the parts of speech are developed according to the requirements of the system. Hence there

¹Karl W. Dykema, "Where Our Grammar Came From," Readings in Applied English Linguistics, H.B. Allen, ed. N.Y., 1958, p.5.

are many more parts of speech within SYSTRAN than the eight or fifteen of human grammars. These parts of speech are established on two criteria: the primary parts of speech are based on congruence of primary function; the secondary parts of speech, on differtiae within the primary groupings.

a. SEGMENT ANALYZER PROGRAM. The Segment Analyzer Program was defined and incorporated in SYSTRAN's parsing apparatus to improve syntactic analysis. The program scans the sentence for specified subordinate clause initiators, coded as such in the stem or idiom dictionary. All words within a clause are marked according to the type of clause initiator. Parenthetic words are also marked. Any words in the sentence which are not marked in this way as belonging either to subordinate clauses or to parenthetic expressions or within parentheses are given main clause indications. Phrases (participial, gerundial, prepositional) are marked before the SEGMENT ANALYZER is activated. They are regarded as parts of the clauses to which their first word is attached.

At each word of the sentence, two bytes are used by the SEGMENT ANALYZER. One contains dictionary information, indicating, for example, that the particular word may initiate a specific type of subordinate clause, or that it may begin a second main clause, or that it belongs to a parenthetic expression or to an idiom. The other byte indicates the level to which the word belongs. An F1 in this byte indicates that the word is part of the first main clause; an F2, that it is part of the second main clause, etc.

The following pages contain lists of (1) subordinate clause initiators, (2) parenthetic words, and (3) words which almost always begin new (second or third) main clauses (i.e., coordinate clauses).

The SEGMENT ANALYZER was designed and implemented to establish a "level" mode of analyzing subsentences, so that lower level (embedded) clauses might be more accurately recognized and handled. It replaced the use of "sentence cuts" in establishing subsentence units. Trained linguists checked SYSTRAN's level output and concluded that level codes do work much more efficiently and accurately than cuts; subsequently, further improvements were made in recognizing levels: rules were established for recognizing colons as main clause initiators; commas were recognized as clause initiators as they simultaneously indicated phrase ends.

SUBORDINATE CLAUSE INITIATORS

<u>Code</u>	<u>Type</u>
01	Qualifying Clauses KAKO1, WHAT KOTORY1, WHICH CE1, WHOSE
02	Objective Clauses KTO (in all cases), THAT CTO (in all cases), THAT
03	Subject Clauses KTO, WHO CTO, WHAT
04	Predicate Clauses KAKOV or KAKO1, WHAT KAKOV or TAKOV, LIKE

<u>Code</u>	<u>Type</u>
05	Clauses of Time KOGDA, WHEN POSLE TOGO, KAK, AFTER PREJDE CEM KAK, BEFORE S TEX POR KAK, SINCE V TO VREM4 KAK, WHILE MEJDU TEM KAK, WHILE POKA, WHILE POKA NE, UNTIL
06	Clauses of Place GDE, WHERE KUDA, WHERE OTKUDA, FROM WHERE
07	Clauses of Comparison and Result KAK, AS KAK BUDTO, AS IF TAK, CTO, SO....THAT TAK CTO, SO THAT
08	Clauses of Cause POTOMU, CTO, BECAUSE OTTOGO, CTO, BECAUSE TAK KAK, BECAUSE, SINCE
09	Clauses of Purpose CTOBY, IN ORDER TO
10	Conditional Clauses ESLI, IF ESLI BY, IF
11	Concessive Clauses XOT4, THOUGH NESMOTR4 NA TO, CTO, IN SPITE OF THE FACT
F1	First Main Clause
F2	Second Main Clause
F3	Third Main Clause
50	Parenthesis

CODE 99: PARENTHETIC WORDS

BESSPORNO	UNQUESTIONABLY
BEZ SOMNENI4	UNDOUBTEDLY
BEZUSLOVNO,	UNDOUBTEDLY
DEISTVITEL6NO,	REALLY
DOLJNO BYT6,	PROBABLY
DOPUSTIM,	LET US ASSUME
ITAK,	THUS
K NAWEMU UDOVOL6STVIH,	TO OUR DELIGHT
K NESCAST6H,	UNFORTUNATELY
K RADOSTI,	FORTUNATELY
K SCAST6H,	FORTUNATELY
K SOJALENIH,	UNFORTUNATELY
K UDIVLENIH,	SURPRISINGLY
K IX UDOVOL6STVIH,	TO THEIR DELIGHT
KAJETS4,	IT SEEMS
KAK VIDNO,	AS IT APPEARS
KAZALOS6,	IT SEEMED
KONECNO,	OF COURSE/CERTAINLY
MOJET BYT6,	PERHAPS/MAYBE
NAKONEQ,	FINALLY/AT LAST
NAOBOROT,	ON THE CONTRARY
NAPRIMER,	FOR INSTANCE
NAVERNO,	PROBABLY
NAVERNOE,	PROBABLY
NESOMNENNO,	UNDOUBTEDLY
OCEVIDNO,	APPARENTLY
ODNAKO,	HOWEVER, BUT
ODNIM SLOVOM,	IN A WORD
PO MNFENIH	IN THE OPINION
PO SLOVAM	ACCORDING TO
PO SLUXAM,	IT IS SAID
PO SU5ESTVU	IN ESSENCE
PO SVEDENI4M,	ACCORDING TO THE INFORMA-
	TION
PO VSE1 VERO4TNOSTI	IN ALL PROBABILITY
POJALU1,	PROBABLY, PERHAPS
PO-MOEMU,	IN MY OPINION
PO-MOEMU MNENIH,	IN MY OPINION
PO-VAWEMU	IN YOUR OPINION
PO-VIDIMOMU	IT SEEMS/IT APPEARS
PREDPOLOJIM,	LET US SUPPOSE
PREJDE VSEGO,	FIRST OF ALL
RAZUMEETS4,	IT GOES WITHOUT SAYING
SLEDOVATEL6NO,	THEREFORE/CONSEQUENTLY
SLOVOM,	IN A WORD
STALO BYT6,	THAT MEANS
TAK SKAZAT6,	IN OTHER WORDS
V CASTNOSTI,	SPECIFICALLY
V KONQE KONQOV,	AFTER ALL
V OSOBENNOSTI,	IN PARTICULAR

V SAMOM DELE,
VERNO,
VERO4TNO,
VIDIMO,
VIDNO,
VQ-PERVYX,
VO-VTORYX,
VPROCEN,
ZNACIT,

IN FACT
PROBABLY
PROBABLY
APPARENTLY/EVIDENTLY
APPARENTLY/EVIDENTLY
IN THE FIRST PLACE
IN THE SECOND PLACE
HOWEVER
THAT MEANS

The above-listed parenthetic words are also available in SYSTRAN's idiom dictionary imbedded within commas.

CODE 98: COORDINATE CLAUSE INITIATORS

I
DA
NI
A
NO
ODNAKO
TO
ILI
LIBO
PO3TOMU
ZATEM

b. REARRANGEMENT. During this optimization phase, several rearrangement procedures were tested, especially those concerning complex sentences. It was found that rearrangement of the subject and predicate in both the main clause and the subordinate

clause resulted in awkward and confusing English sentences, since the predicates of both clauses were juxtaposed at the end of the sentence. It was also discovered that, if in the Russian sentence the predicate precedes the subject, then rearrangement should be effected only if in the translation the subject and predicate are not separated by twelve or more words.

Special subroutines have been incorporated into the main program to improve SYSTRAN's rearrangement technique. These subroutines carry out range searches of the items to be rearranged. The predicate range search encompasses all objects (direct and indirect) of the verb and all of its adverbial modifiers. These items are marked so that they may be moved with the verb. A double indexing technique makes it possible to rearrange items within the specified range, as well as to move all the items specified by the range as a unit. (Adverbs which should not be rearranged carry codes to that effect.)

Similarly, the noun range routine checks for all modifiers of the subject(s). Further routines were introduced to determine the proper governors of prepositional phrases occurring between constituents marked for rearrangement. Additional routines were required for names, English letters, mathematical formulae, etc., to determine their proper point of attachment before rearrangement was effected.

c. A COMPARISON OF HUMAN AND SYSTRAN PARSING. The following sentence has been chosen to illustrate SYSTRAN's present parsing capabilities by comparing two grammatical analyses of

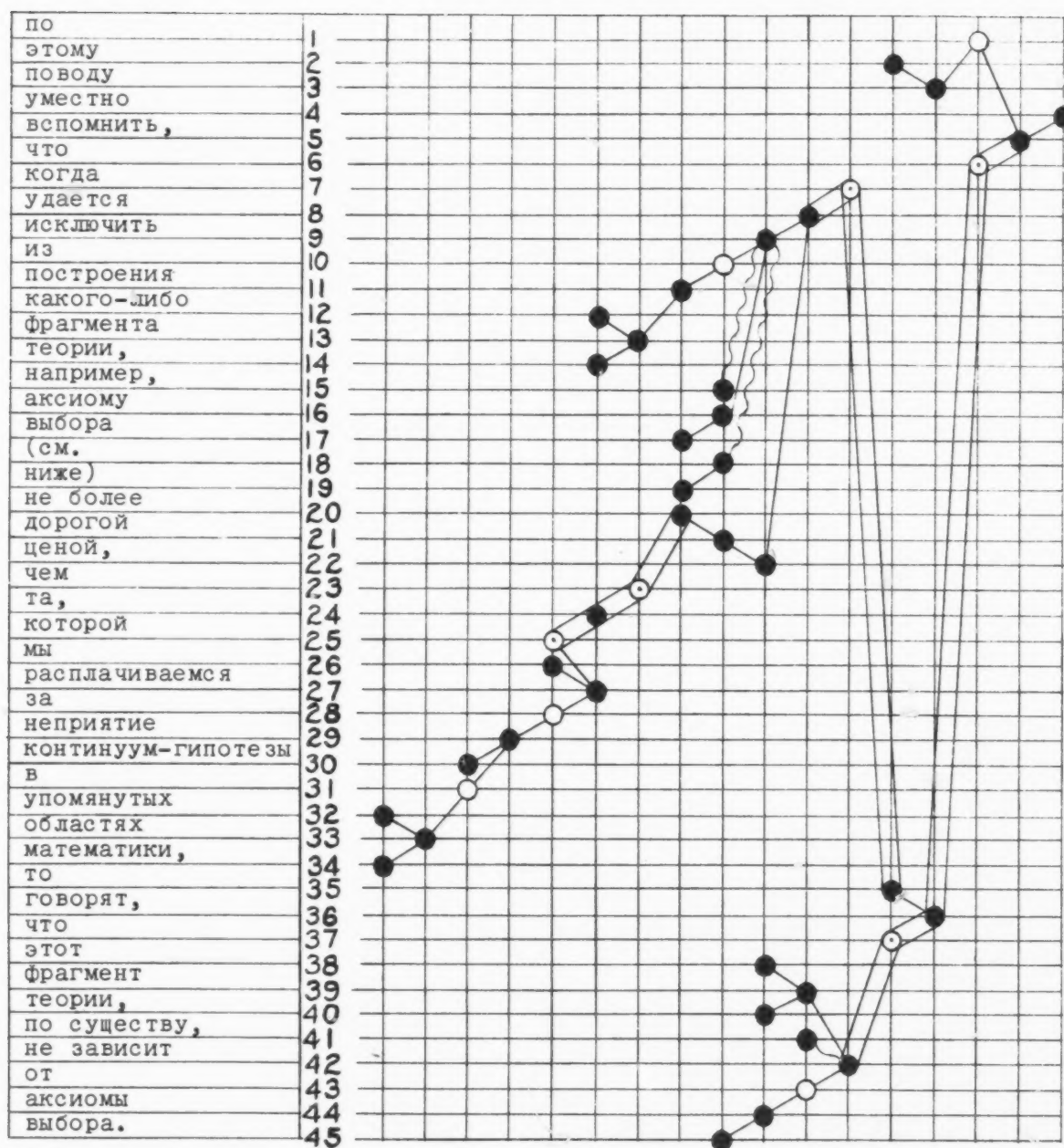
it: one done by Soviet linguists in Russia; the other, by SYSTRAN.

The sentence, from Sevbo, I. P., "O gromozdkosti sintaksičeskix struktur," Naučno-Texničeskaja Informacija: Avtomatizacija perevoda tekstov, ser. 2, No. 2, Moscow (1971), p. 42, is the following:

PO 3TOMU POVODU UMEŠTNO VSPOMNIT6, ČTO KOGDA UDAETS4
ISKLHCIT6 IZ POSTROENI4 KAKOGO-LIBO FRAGMENTA TEORII,
NAPRIMER, AKSIOMU VYBORA (SM. NIJE) NE BOLEE DOROGO1
QENOL, CEM TA, KOTOROL MY RASPLACIVAEMS4 ZA NEPRI4TIE
KONTINUUM-GIPOTEZY V UPOM4NUTYX OBLAST4X MATEMATIKI, TO
GOVOR4T, ČTO 3TOT FRAGMENT TEORII, PO SUŠESTVU, NE ZAVISIT
OT AKSIOMY VYBORA.

This Russian sentence was analyzed by SYSTRAN's parsing program and codes were attached accordingly. It is not the aim of this report to enumerate and elaborate upon all the codes developed and used within SYSTRAN. Rather, the aim is to show the results of human and mechanical parsings of the same sentence. Figure 1 shows the human analysis; Figure 2, the SYSTRAN analysis. This type of diagram is used here only because it is the type used in the Soviet article. Unfortunately, the graph does not allow many SYSTRAN analysis codes to be shown. For example, while SYSTRAN points from a governor to a governed word, and back to the governor from the governed word, that cannot be shown here.

The numbers and letters along the right-hand side of Figure 2 indicate SYSTRAN clause levels: F1 identifies the first main clause (F2 would indicate a second main clause); 02 identifies a CTO clause; 50, words with parenthesis; 01, a



- FULL MEANING WORDS
- ⊙ CONJUNCTIONS & CONNECTIVE WORDS
- PREPOSITIONS
- SUBORDINATE CLAUSES
- GOVERNMENT RELATIONSHIPS
- ~ "ISOLATED" WORDS

FIGURE 1

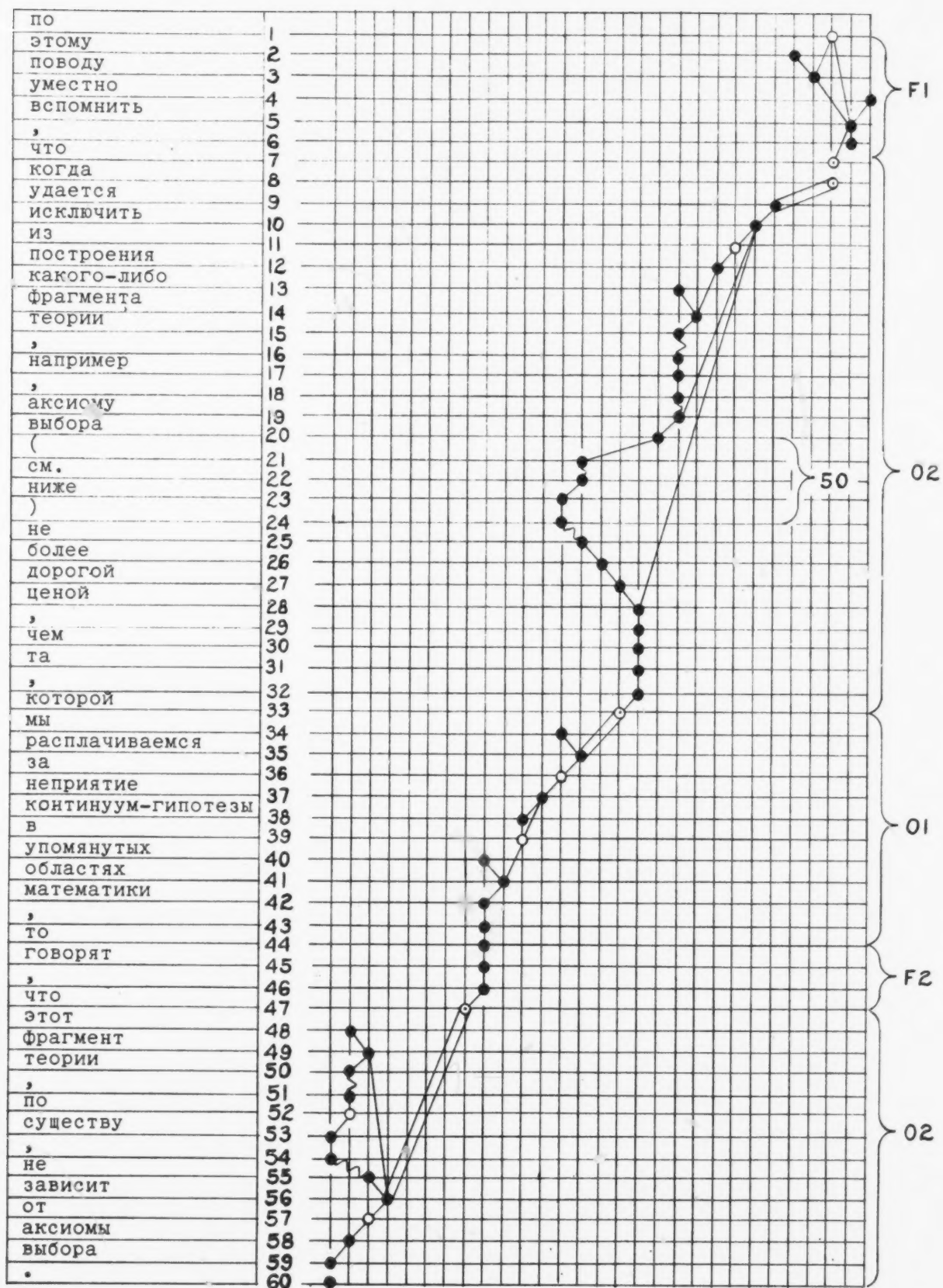


FIGURE 2

KOTOROl clause; etc. (See list in I.2.a.)

The major difference in the illustrations is the attachment of the first CTO to the rest of the sentence. The human analysis attaches it directly to GOVOR4T, which is linked to TO which, in turn, is linked to KOGDA. TO is attached to nothing except KOGDA and GOVOR4T. This reflects the major difference between the human translator and the computer: the human can intuitively infer the author's intention, while the computer must rely on what the author has actually written.

In this particular instance, the author, ignoring traditional rules of Russian grammar, has omitted the comma between CTO and KOGDA. (Normally, any new subordinate clause marker will require a comma.) The human analyst "reads in" the comma, thereby creating or recognizing that the KOGDA clause is embedded in the CTO clause. The computer, meeting this unusual circumstance of adjacent subordinate clause initiators which are not separated by a comma, has no indication that two subordinate clauses are initiated virtually simultaneously. Therefore (Figure 2), it assigns a CTO clause marking to CTO, to KOGDA and to the entire ensuing clause.

It will be seen, however, in the following, that the translated output, i.e., the actual translation done by SYSTRAN, was not affected by this failure to identify the KOGDA clause separately from the CTO clause:

IN REGARD TO THIS IT IS APPROPRIATE TO RECALL THAT WHEN IT IS POSSIBLE TO EXCLUDE FROM THE CONSTRUCTION OF ANY FRAGMENT OF A THEORY, FOR EXAMPLE, THE AXIOM OF CHOICE (SEE BELOW) NOT BY A MORE EXPENSIVE PRICE THAN THAT BY

WHICH WE PAY FOR THE NONACCEPTANCE OF A CONTINUUM HYPOTHESIS IN THE MENTIONED REGIONS OF MATHEMATICS, THEN IT IS SAID THAT THIS FRAGMENT OF A THEORY, IN ESSENCE, DOES NOT DEPEND UPON THE AXIOM OF CHOICE.

In order to further demonstrate SYSTRAN's parsing abilities, another sentence, analyzed by L. N. Iordanskaja,² is illustrated, along with SYSTRAN's analysis of the same sentence. Again, the codes used by SYSTRAN have been "translated" into a format compatible with that chosen by Iordanskaja. In her diagram, she used arrows to point from governors to words they govern. It should be kept in mind that SYSTRAN's hexadecimal output also shows pointers from governed words back to their governors. Figure 3 shows Iordanskaja's analysis; Figure 4, SYSTRAN's analysis.

In this particular instance, SYSTRAN's analysis was more detailed than the human one, in that it shows the connections between, as well as within, clauses. Thus, KOTOROE points back to TREBOVANIE.

Moreover, the human analysis leaves the conjunction A dangling. It is not shown to be attached to any other word in the sentence. SYSTRAN, however, recognizes that both KLADOISKATELEM and UCENYM are complements of BYT6. It recognizes this relationship by recognizing a relationship between NE and A: if a comma and A (,A) are preceded by NE and a noun or adjective, and if no

²L. N. Iordanskaja, "Automatičeskij sintaksičeskij analiz," Kibernetika v monografijax, 2, izd. Nauka, Novosibirsk (1967), p. 49-74.

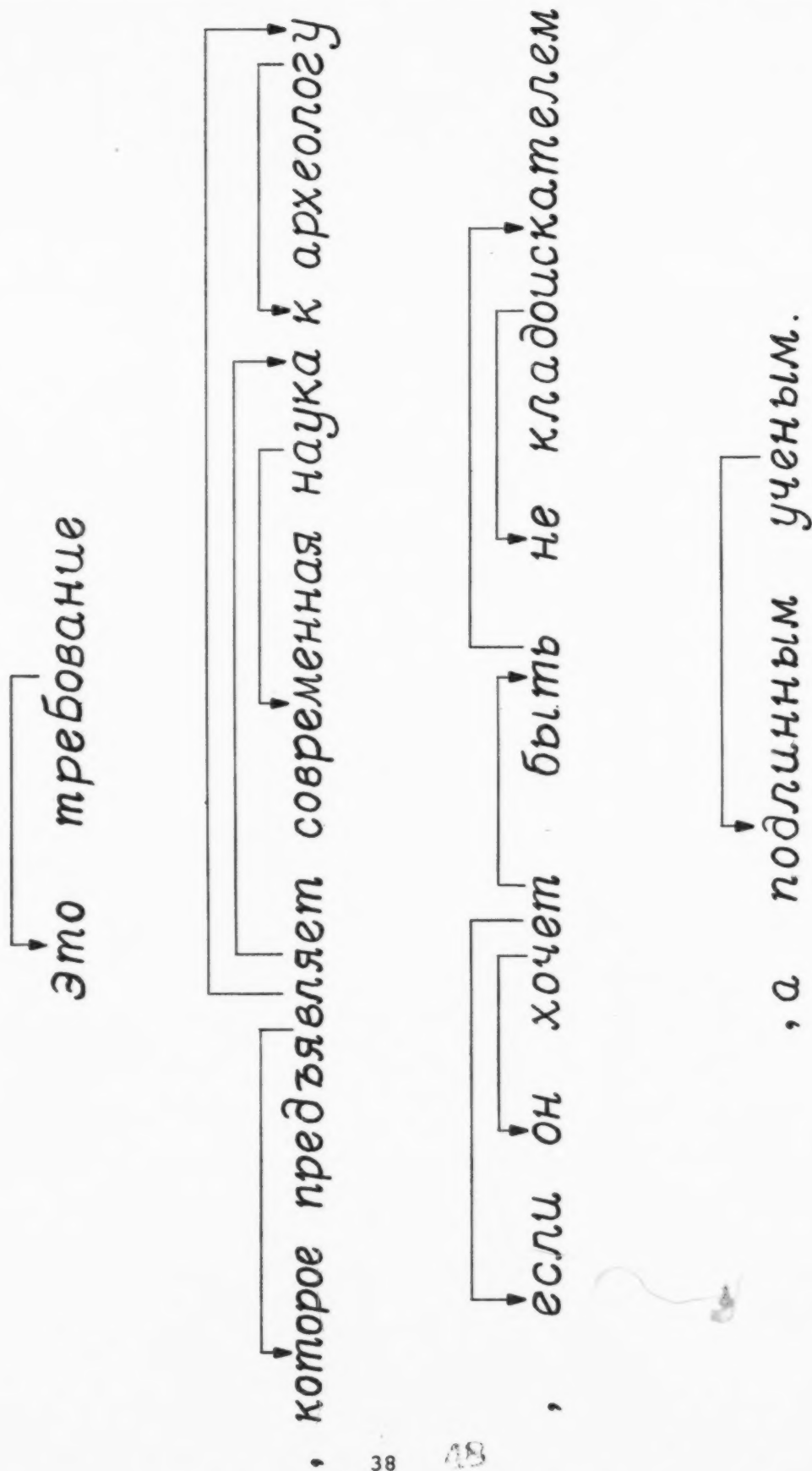


FIGURE 3

JORDANSKAJA'S ANALYSIS

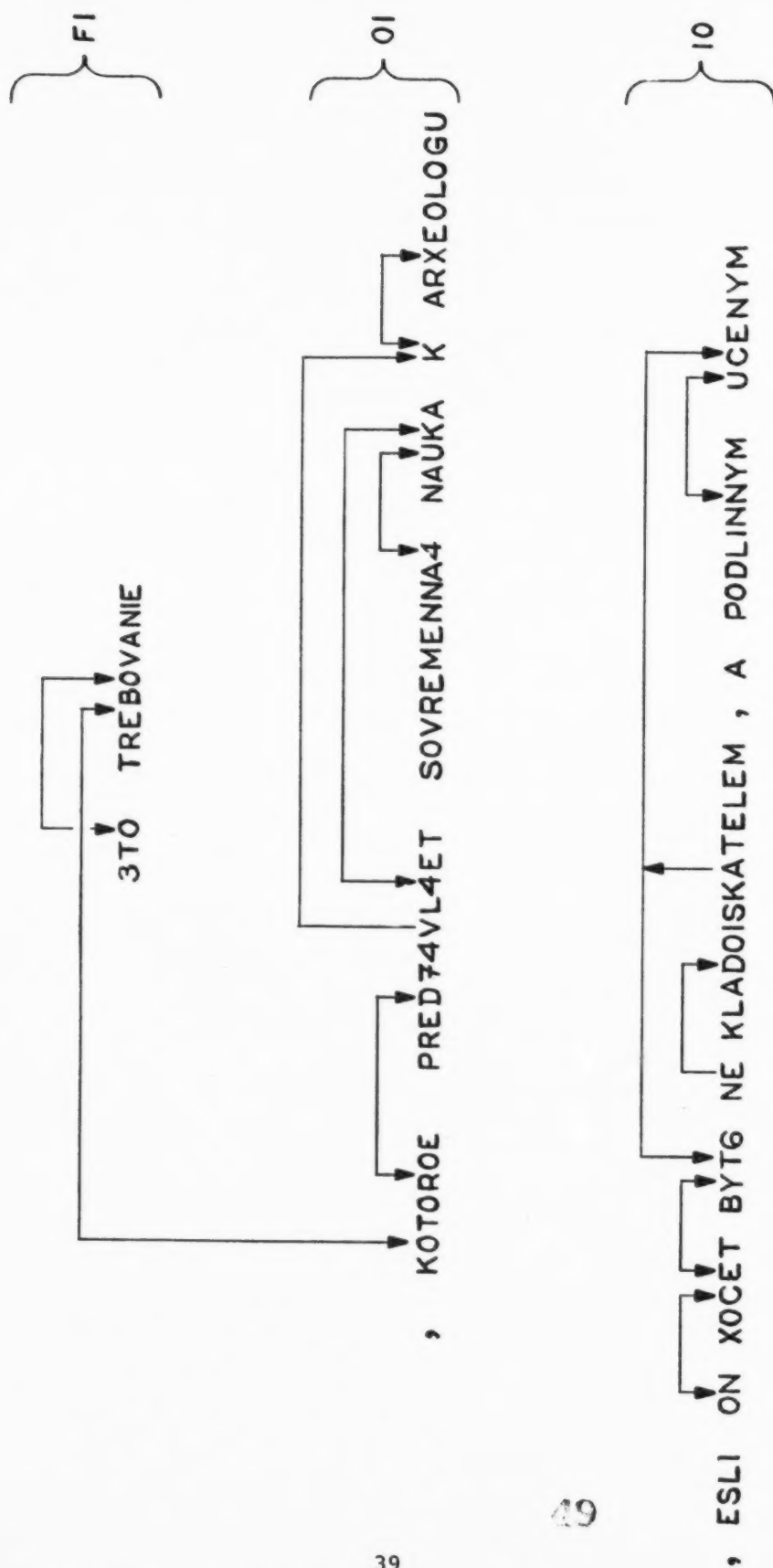


FIGURE 4
SYSTRAN ANALYSIS OF IORDANSKAJA'S SENTENCE

predicate is found in the A phrase, then the nouns or adjectives following NE and A serve parallel functions. The incorporation of this grammatical principle also allowed SYSTRAN to differentiate between UCENYM as a dative plural and UCENYM as an instrumental singular. Therefore, the following translation was achieved:

THIS IS THE REQUIREMENT WHICH CONTEMPORARY SCIENCE TO AN
ARCHAEOLOGIST PRESENTS, IF HE WANTS TO BE NOT A HUNTER,
BUT AN AUTHENTIC SCIENTIST.

4. SEMANTICS

For computer systems such as SYSTRAN, it is best to regard the meaning of any lexical item as an aspect of performance and to class certain lexical forms as items characterized by exponential features known in the outside language as primary dimension. That is, the meaning of a word varies within certain definable limits according to associative relations built within sets of circumscribed syntagmatic variables.

The task from which improvements related to semantics are to emanate consists primarily of selecting polysemantic words establishing existing syntagmatic relations, and devising a working procedure which automatically narrows the wide divergence of relationships regularly occurring in Russian technological texts by establishing valid generalized rules.

The available evidence shows, for instance, that the word KOLICESTVO functions with several meanings, two of which possess distinctive denotations, each depending on the semantic category characterizing the word with which KOLICESTVO is grammatically and syntagmatically associated. In the translation, either of two different meanings is assigned: AMOUNT or NUMBER.

AMOUNT denotes an aggregate on the sum total of two or more quantities and KOLICESTVO carries this meaning when it forms a grammatical unit with words that belong to the semantic category comprised of items designated as collective or non-unit nouns. These are SOL6, MASLO, PAR, VODA, NEFT6, TEMPERATURA, TEPL0, BIOMASSA, KISLOROD, SAXARY, PRIMES6, POROWOK, TALLI1, ORTOFOSFAT, PLATINA, KALI1, GELI1, MED6, KREMNI1, etc.

NUMBER, on the other hand, denotes the sum total of a collection of units and KOLICESTVO carries this meaning when it forms a grammatical unit with words that belong to the semantic category comprised of items designated as countable or unit nouns. These are IZMERENIE, PLENA, TOCKA, ZERNO, INGIBATOR, VARIANT, OB7EKT, PREDKAZANIE, MIKROORGANIZM, BAKTERI4, MOLEKULA, FERMENT, SPORA, CLEN, MIKROB, MUTANT, QIKL, ZAMYKANIE, SLO1, etc.

Although the above-listed nouns may share several semantic categories, the two mentioned here suffice in order to write a program which clearly distinguishes between the two meanings and selects the appropriate one during the structural analysis of KOLICESTVO, provided all pertinent codes are attached to both

KOLICESTVO and the words in the list.

As an example, two sentences are given here to illustrate SYSTRAN's capabilities in carrying out improvements on the level of semantics.

V TABL. 3 PRIVEDENY REZUL6TATY 3KSPERIMENTOV, V KOTORYX
IZUCALOS6 VLI4NIE SILY TOKA I KOLICESTVA JELEZNOGO POROWKA
V POKRYTII.

In this sentence KOLICESTVA is grammatically tied to POROWKA. This activates a program which selects the meaning AMOUNT.

DL4 KOLICESTVENNOGO OPREDELENI4 FAZ PRIMENEN TOCECNYI
METOD, KOTORYI SVODITS4 K PODSCETU KOLICESTVA TOCEK,
PRIKOD45IXS4 NA KAJDUH FAZU IZ OB5EGO KOLICESTVA ZEREN.

In this sentence KOLICESTVA is grammatically tied to TOCEK and ZEREN. This activates a program which selects the meaning NUMBER.

SECTION III

FURTHER IMPROVEMENTS EFFECTED DURING OPTIMIZATION

1. INTRODUCTION

In order to demonstrate SYSTRAN capabilities in translating Russian sentences into English, live data were used to substantiate each discussion in this section of the report. During output analyses, sentences were selected on the basis of the particular areas covered by the Work Statement in order that they could be used as practical examples illustrating the appropriate improvements. Homograph examples are shown at the end of this section in reductions of actual computer print-outs.

a. IMPROVING THE TRANSLATION OF HOMOGRAPHS. Russian words identified as homographs (lexical or morphological) are entered into the Stem Master Dictionary as full-form entries, each identical form carrying its appropriate grammar codes (i.e., both morphological and syntactic information) and the corresponding meaning with its synthesis codes. The full-form entries comprising a homograph are differentiated in the dictionary by a sequential number, or, a Dictionary Stem Number (DSN). DRUGOM, for instance is entered under DSNOO and DSN01, and the codes indicate that the '00' entry is a noun, masculine singular, instrumental, animate with the meaning FRIEND while '01' is an adjective (actually, a noun-adjective ambiguity), masculine, neuter, singular, locative meaning OTHER. The numbers '00,' '01,' '02,' are assigned to the different entries of the

homograph according to descending order of predicted use. All entries identified as homographs carry the necessary code HM (homograph).

During translation processing within the Main Dictionary look-up pass, a textword that is in the dictionary as an HM entry is matched with the entry having DSN00. A normal output processing record is built containing the associated grammar and target language meaning. At this time multiple output processing records are prepared for each of the additional homograph entries in the dictionary.

Prior to parsing, a standard analysis area is built for each textword in the sentence. This area provides for a maximum of 105 words per sentence. Following this area is a provision for an additional 30 analysis word areas where the multiple homograph word entries are placed. Each word entry in the standard area that is identified as a homograph entry contains in a byte the relative word sequence in the combined analysis area of the first additional homograph entry for that word. Then each additional homograph entry contains the sequence number of the next until the last which has a zero sequence number.

In Structural Pass Zero (STRPASS0) the standard analysis area is scanned for homograph indications and routines are activated by routine numbers carried as dictionary codes in the stem. Each routine examines the environment of the word in the sentence and a decision is made as to the desired part of speech

for the function of the homograph. Using standard processing MACROS, the homograph record for this part of speech is located and this analysis word entry is moved to the location in the standard area. It may be that the desired entry is the one already in the area, in which case no movement takes place. Each successive word is scanned for a homograph indication until the end of the sentence is reached. The standard analysis area contains, at this point, the final grammar and target language translation pointers to be used throughout the translation process for that sentence.

For the purpose of avoiding discrepancies that might result from the presence of homograph-type words, selected groups were established on the basis of which homographs can be properly dealt with at specific points during the translation processes. These groups are given below, containing appropriate examples and a brief description of the solution of the homograph ambiguity.

PREPADR

Resolves Preposition/Adverb Ambiguity

HM: MIMO	SVERXU
NAPROTIV	SZADI
POPEREK	VBLIZI
POSLE	VDOL6
POSREDI	VNUTRI
POVERX	VOKRUG
POZADI	VOZLE
PREJDE	VPEREDI
SBOKU	VSLED

PREPADR scans to the right from the homograph for three words, searching for a word in the genitive. If the genitive is not found, the preposition-homograph is replaced by the adverb-homograph.

PROFR

Resolves Profession/Title Ambiguity

HM: AKUSTIK/A	BIOMEXANIK/A
ANALITIK/A	BOTANIK/A
ASTROBIONIK/A	GENETIK/A
ASTROFIZIK/A	MATEMATIK/A
AVIAMEXANIK/A	RENTGENOTEXNIK/A
AVIATEXNIK/A	SEMIOTIK/A
A3ROMEXANIK/A	TERMO3NERGETIK/A
BIOFIZIK/A	TEXNIK/A
BIOMATEMATIK/A	

This routine selects the title-homograph when a proper name follows the current word within three words or the current word is modified by ZNAMENIT or IZVESTN, or the governor or the modifier to the left resolves the ambiguity grammatically. It selects the profession-homograph when the current word is governed by PO, PRO, or KASATEL6NO, or the governor or the modifier to the left resolves the ambiguity grammatically.

DRUG

Resolves Noun/Adjective Ambiguity

HM: DRUGOM
BELOM (in acoustics)
PRAVOM

DRUG scans four words to the left of the current word in search for a vehicle-word requiring the locative case. If this condition is satisfied, the adjective-HM is used. If not, the noun-HM is selected.

NACR

Resolves Noun/Verb Ambiguity

HM: NACALO
NACALA

This routine selects the verb-HM if the current word is followed by an infinitive or if there is neither a finite verb nor an auxiliary, nor a short form neuter singular in the sub-sentence. It also selects the verb when a neuter finite verb or an auxiliary is encountered which is enumerated with the current word.

EST6R

Resolves EAT/Present Tense Auxiliary Ambiguity

HM: EST6

This routine searches three words to the left of the current word for a word with an infinitive requirement. If found, the infinitive-HM EAT is used, otherwise the finite verb form, present tense, singular or plural, of the auxiliary BE is used.

INADR

Resolves Instrumental/Adverb Ambiguity

HM: CAST6H	POLNOST6H
DNEM	QELIKOM
KRUGOM	R4DOM
LETOM	SKACKOM
NOC6H	UTROM
OSEN6H	VECEROM

This routine asks if the word to the left of the current word is an adverb or 'NE'; if so, the adverb is used. If the word after the current word is a verb or short form adjective, the adverb is used again. If a verb follows the current word and requires the instrumental case, the adverb homograph is used. If the word after the current word is an adjective, adjectival noun, pronoun adjective, or a participle and the word before the current word is in the instrumental singular, the noun homograph is used. If there is an instrumental governor to the left, the noun is used. If the following word is a noun (pure noun, proper noun, or profession) forming an adnominal genitive construction, the noun homograph is used.

VPARTR

Resolves Short Form Present Passive Participle

Masculine/First Person Plural Present Indicative

HM: DOPUSTIM	SOKRA5AEM
IZBIVAEM	SORBIRUEM
IZGIBAEM	SRAVNIVAEM
IZLECIM	STEKLUEM
MODELIRUEM	STR4SAEM
OB74SNIM	SUMMIRUEM

OTDAVAEM	SVODIM
OTVODIM	TERPIM
O5UTIM	UPROCN4EM
O5U5AEM	VIDIM
PERERASTAEM	VMATYVAEM
PERETACIVAEM	VMEN4EM
PEREVODIM	VNOSIM
POLOJIM	VOOBRAJAEM
POSYLAEM	VOZBUDIM
POZNAVAEM	VOZGON4EM
PRINIMAEM	VOZGORAEM
PRIPISYVAEM	VYCISL4EM
PRIVODIM	VYSTAVL4EM
PRODOLJAEM	VY4SNIM
PROIZVODIM	ZABIRAEM
PROXODIM	ZABYVAEM
RASPOLAGAEM	ZADUVAEM
RASSMATRIVAEM	ZAKARMLIVAEM
RAZDELIM	ZAMEN4EM
RAZLAGAEM	ZAMETAEM
REGULIRUEM	ZAMORAJIVAEM
SGORAEM	ZANIMAEM
SJIMAEM	ZAVISIM
SKLON4EM	ZNACIM
SMEN4EM	ZRIM
SMEWIVAEM	3KSTRAGIRUEM
SOEDINIM	

This routine checks the word before the current word. If a finite auxiliary, the short form participle homograph is used. It is also used if the scan to the left locates ON or a nominative singular case noun. The verb is used if MY is found in the left scan. It is also used when MY is to the right, provided no adjective or accusative noun, or participle is to the left.

NEUTFM

Resolves Neuter/Feminine Ambiguity

HM: JELEZO
JELEZA

NEUTFM determines if the current word is preceded by an adjective, then it is the adjective which, on the basis of grammatical agreement determines the correct choice. For the resolution of the ambiguous -A ending, the neuter-HM is chosen when a noun immediately precedes since this suggests an adnominal genitive construction. The ambiguous -U, -E endings present a need for a further search within the sentence for governors which resolve this ambiguity.

VESR

Resolves Verb/Noun Ambiguity

HM: VESTI

This routine selects the infinitive if the word preceding the current word has the infinitive requirement. In all other cases, the noun is used.

SUTR

Resolves Verb/Noun Ambiguity

HM: SUT6

This routine checks simultaneously to the left and to the right of SUT6 for a nominative plural noun. If both are found, the verb homograph is used. In all other cases, the noun homograph is set in the translation.

NPREPR

Resolves Noun/Preposition Ambiguity

HM: POD

This routine looks to the right of the current word for a noun phrase [i.e., (adverb) + (adjective) + noun] in the instrumental case. If such a noun phrase is not found, the noun homograph (meaning and codes) replaces the more common prepositional entry.

UJER

Resolves Comparative/Adverb Ambiguity

HM: UJE

This routine checks on the word following the current word. If it is a common noun, proper noun, or a noun denoting a profession and it is in the genitive case and if the word preceding the current word is an adjective, adjectival noun, or participle also in the genitive case, the adverb is chosen. If the preceding word is not in the genitive case, the comparative is chosen.

CONJAR

Resolves Conjunction/Abbreviation Ambiguity

HM: A

This routine sets the current word to the abbreviation homograph if the word is capitalized and does not begin the

sentence. The abbreviation HM is also selected if a comma follows, which is in turn followed by the letter 'B,' a period, or closed parenthesis. If a comma follows the current word (and there is no 'B' after it), the conjunction is chosen.

FIMNAR

Resolves Feminine Instrumental Singular/Masculine
Nominative-Accusative Singular Ambiguity

HM: PROBO1

This routine checks the word immediately to the left of the current word. If it requires the instrumental (i.e., if it is 'S,' 'NAD,' 'POD,' or 'PERED'), the feminine noun is chosen. If an adjective or participle is found immediately preceding the current word, it is checked for case. If nominative, the masculine noun is used, and if instrumental, the feminine noun is used.

NCOMPR

Resolves Noun Feminine Dative-Locative/Comparative
Ambiguity

HM: GU5E

This routine scans to the right for the word 'CEM' and, if it is found, the comparative homograph is chosen. If an adjective or participle is found immediately to the left or in a scan to the left, and if it is locative or dative and feminine

singular, then the feminine noun homograph is chosen. If the last two conditions are not met, the comparative is selected.

VSEMR

Resolves Locative-Dative/Instrumental Ambiguity

HM: VSEM

This routine checks the word in front of the current word. If it is one of the prepositions 'V,' 'NA,' 'O/OB,' or 'PRI,' the locative/dative form of 'VSEM' is chosen. All other cases point to the instrumental form of the word.

ABVBR

Resolves Imperative/Abbreviation Ambiguity

HM: SM.

This routine checks if the current word is the sentence beginning. If it is, the imperative homograph is selected. If a numeral directly precedes the current word, the abbreviation is chosen. In all other cases, the imperative form is used.

KOMR

Resolves Noun/Pronoun Ambiguity

HM: KOMU
KOM

If the word encountered is KOMU, and there is an accusative, adjective or participle immediately to the left, the noun

homograph is used. If the adjective or participle is in any other case, the pronoun is selected. If no adjective or participle is encountered, there is a scan to the right and the left for a word requiring the dative or a word requiring the accusative. If the dative requirement is found, the pronoun is used, while the noun homograph is chosen if an accusative requirement is found.

If KOM is the current word, and the word immediately to its left requires the locative case, the pronoun is used. If none exists, the noun homograph is chosen.

SPOR

Resolves Masculine/Feminine Ambiguity

HM: SPOR/A

This routine scans to the right of the current word to the end of sub-sentence unit for O or OB. If found, the masculine 'SPOR' homograph is selected. If a genitive plural follows the current word, it is made the feminine 'SPORA.' If the word before the current word (an adjective) ends in -l and the current word has a null ending, the masculine homograph is chosen. If the adjective ends in -YM, again the masculine is chosen. If the current word ends in -OM, the same choice is made. If the current word ends in -Ol, the feminine homograph is used. More endings are checked, and the appropriate homograph is assigned.

SUBJECT CODE C.. (PREPAR) IN 1954 DURING JUNE THEY CAUGHT ORGANISMS NEAR THE POPULATED POINTS.

(PREPAR) THIS LEADS TO AN INCREASE IN THE NUMBER OF TRANSMISSIBLE INFECTIONS IN THE LOCATED NEAR VILLAGES.

(PREPAR) THE CONCEPT OF UNIFORMITY IN THIS CASE IS EXTENDED BOTH TO THE SEMANTICS AND TO STATISTICS, OR TO THE TWO TOGETHER.

(DRUG) HOWEVER THE AUTHOR FORGOT ABOUT PAVLOV'S ANOTHER KNOWN INDICATION.

(DRUG) THIS SCIENTIST IS THE PERSONAL FRIEND OF MANY SOVIET SCIENTISTS.

(NACR) AFTER SUNSET IT BEGAN RAPIDLY TO GROW DARK.

(NACR) IN KINDERGARTEN, WHERE THE BEGINNING OF AN OUTBREAK WAS POSSIBLE TO REVEAL, QUANTITIES CONTINUED 30 DAYS.

(EST6) A GREAT TEACHER LENIN TEACHES ALL, THAT THE SOVIET CITIZEN SHOULD EAT A NUTRITIOUS FOOD.

(EST6) THIS RESULT THE MAIN SUPPORT OF CAREFUL OPTIMISM IS.

(INADP) ESPECIALLY AT NIGHT STATE WAS MADE WORSE.

66

(INADP) WERE ALWAYS, EVEN IN CALM WEATHER, SOMETHING RUZZED IN OLD AMASOV FURNACES, AND DURING A THUNDER STORM THE ENTIRE HOUSE SHOOK AND IT SEEMED IT WAS CRACKED APART, AND IT WAS TERRIBLE, ESPECIALLY AT NIGHT, WHEN ALL TEN LARGE WINDOWS WERE SUDDENLY ILLUMINATED WITH LIGHTNING.

54B

(INADP) THE EXTRACTION OF ACIDS BY THE MECHANISM FOR A CONNECTION HAS BEEN CAUSED BY THE FORMATION OF THE CONNECTION BETWEEN THE ELECTROPHILIC PART OF THE MOLECULE OF AN ACID AND ELECTRON-DONOR GROUP OF THE MOLECULE OF AN EXTRACTANT 2.

(VPAPT) A VIRUS WAS INSERTED BY THEM INTO AN ORGANISM.

(VP22) WE INSERT THE APPROPRIATE ADDITIONS INTO A PROGRAM.

(NEUM) DYPHTHERIA BACTERIA IN A MEDIUM WITH THE SCARCITY OF IRON CAN
SYNTHESIZE AN ALBUMINOUS PART.

(VES) THE ACCOUNT OF THIS QUESTION IN A TEXTBOOK CANNOT BE CONDUCTED
SEPARATELY FROM THE DATA OF THE PRACTICE OF PUBLIC HEALTH.

(VES) A FAMOUS ACTRESS DISAPPEARED WITHOUT NEWS.

(SUT) THE ESSENCE OF THIS METHOD IS INCLUDED IN THE FOLLOWING.

(NPRE2) UNDER HIS MANAGEMENT MANY SPECIALISTS GREW.

(UJER) THE RIGHT OPENING IS NARROWER THAN A LEFT OPENING.

(UJER) THE GENERALITY OF THE PHENOMENA OF PHAGOCYTOSIS WAS AT THIS TIME

ALREADY WELL KNOWN.

(CONJA) BUT IN THESE GROUPS WAS EQUAL A AND A, BUT IN CONTROL - A.

(VSEFMR) THE STATE OF OUR KNOWLEDGE ON ALL THESE QUESTIONS IS LOCATED IN THE
STAGE OF HYPOTHESES.

(VSEFMR) KHRUSHCHEV RELATED TO THE PATIENTS WITH ALL ATTENTION, WHICH THE
POSITION OF A FATHER REQUIRED.

(ARVP) DEVIATIONS FROM THE UNITS OBTAINED DIRECTLY AS A RESULT OF THE
CALCULATION, DO NOT EXCEED 55 (SEE THE SEVENTH ON A VERTICAL LINE TO THE
GRAPH OF TABLE).

(ARVP) THEY ALWAYS LOADED AN ELECTRODE INTO A MEDIUM TO THE DEPTH IN 5 CM..

b. ADVERBS. In the effort to improve SYSTRAN's translation of adverbs, work was carried out using live data from which sentences containing adverbs were extracted and sorted according to the nature of needed improvements. Certain adverbs were selected and, on the basis of existing SYSTRAN rules, methods were sought by which additional or new rules could be devised and, upon being tested during fresh computer runs, incorporated into the System. A special subroutine designed to deal with those single-word Russian adverbs which are best rendered into English as adverbial phrases comprised of prepositions and nouns is described below. The prepositions function normally (as synsemantic words), while the nouns are the carriers of the meaning.

A representative example of an adverb of this sort is the word VNACALE which has the meaning IN THE BEGINNING when most appropriately translated into English. When this adverb is concorded with a word functioning as predicate (or, which itself is concorded to the left with an auxiliary), and the natural word order is maintained, the translation is stylistically awkward:

. . . VNACALE NE VOZNIKAET . . .

. . . DOES NOT IN THE BEGINNING ARISE . . .

This substantiates the new ESYN modifications. The "adverb" or "POS 30" routine checks the English area of the Dictionary Meaning and checks for the words AT, BY, FOR, FROM, IN, and TO

followed by a space and another word. If these conditions are met, the adverb is rearranged behind the predicate and the concord between that adverb and verb is broken. Other adverbs, should they precede the verb, remain intact for the new ESYN inserts. Thus the translation becomes as follows:

. . . VNACALE NE VOZNIKAET . . .
. . . DOES NOT ARISE IN THE BEGINNING . . .

Should the sentence be modified by insertion of additional adverbs located before the finite verb, the translation would be similar to the following example:

. . . VNACALE NE OCEN6 CASTO VOZNIKAET . . .
. . . DOES NOT VERY OFTEN ARISE IN THE BEGINNING . . .

Other one-word Russian adverbs actually used in scientific texts have been encountered and are now treated according to the rules of this POS 30 Routine.

Examples are given below:

VPERVYE
FOR THE FIRST TIME

DETAL6NO
IN DETAIL

NADOLGO
FOR A LONG TIME

Additional adverbs falling into this group can be compiled as they are found in the texts in order that they also may be subjected to the rules of this routine, and contribute to the amelioration of the translation procedure.

Further modification to the ESYN programs affecting POS 04 (finite verbs) and POS 44/48 (infinitives) has been accomplished. The new rule splits the inserted auxiliary material before the first adverb concorded with the verb. The formula is as follows:

Aux + Adv + Verb (finite or infinitive)

In case of an infinitive, the TO, having been suppressed at the verb, is inserted before the adverb. The formula is:

TO + Adv + Verb (infinitive).

In case of finite verbs, only the main verb form and, for reflexive verbs, the reflexive pronoun are inserted at the verb while the auxiliaries, pronouns, and negations are inserted before the adverb. The formula is

Aux/Pron/Neg + Adv + Verb (finite)

The procedure described in this rule will ensure that the collocated expression ODN OVREMENNO VY4VL4ETS4 is translated IS SIMULTANEOUSLY EXPOSED.

Special "adverb" routines based on an adverb as the current words are being developed for the purpose of improving the translation procedures. Two such lexical routines whose current

words are XOROWO and SOVSEM are given here.

The XOROWR is initiated by the text word XOROWO which carries the dictionary meaning WELL. This translation is modified to GOOD when the word functions as a predicate or when it is not concorded to right or left.

Example: XOROWO, CTO V OCERKAX, POSV45ENNYX OTDEL6NYM
INFEEKIAM . . .

IT IS GOOD THAT IN DESCRIPTIONS, DEVOTED TO
SEPARATE INFECTIONS . . .

When XOROWO is concorded to the right or left, it is translated WELL.

Example: 3TI REZUL6TATY XOROWO SOGLASUHTS4 S DANNYMI
DRUGIX ISSLEDOVATELEL.

THESE RESULTS WELL AGREE WITH THE DATA OF
OTHER INVESTIGATORS.

or: INFEEKI OPISANY XOROWO.

THE INFECTIONS ARE DESCRIBED WELL.

The SOVSEM Lexical Routine is initiated by the text word SOVSEM which carries the dictionary meaning ENTIRELY. This translation is modified by preceding and following verbal elements.

If SOVSEM is preceded by a negated finite verb, negated auxiliary, or negated infinitive, the translation is made AT ALL.

Example: ONI NE AGGLHTINIROVALIS6 SOVSEM V BOLEE POZDNE
SROKI.

THEY WERE NOT AGGLUTINIZED AT ALL IN LATER
PERIODS.

SOVSEM preceded by a non-negated verbal is rendered ENTIRELY
from the SYSTRAN Dictionary.

Example: ONI AGGLHTINIROVALIS6 SOVSEM.

THEY AGGLUTINATED ENTIRELY.

If SOVSEM is followed by the negating particle NE and a
correctly negated word, this negation is dropped and the trans-
lation is modified in one of two ways, depending on whether or
not MALO ILI are in the same clause. If SOVSEM is preceded by
MALO ILI, the meaning IF is assigned to ILI and the meaning AT
ALL to SOVSEM.

Example: MIKROBY MALO ILI SOVSEM NE VIRULENTNY.

THE MICROBES ARE LITTLE IF AT ALL VIRULENT.

If neither MALO nor ILI are there, the meaning IN NO WAY is
assigned to SOVSEM.

Example: 3TOT METOD 4VL4ETS4 METODOM, KOTORY1 PRIMEN4ETS4
ILI SOVSEM NE PRIMEN4ETS4.

THIS METHOD IS THE METHOD WHICH IS APPLIED OR IS
IN NO WAY APPLIED.

Additional adverbs were selected and special "adverb"
routines were developed for them. Three such lexical routines

whose current words are DAVNO, OBRATNO, and VES6MA are also given here.

DVNOR

This routine is initiated by the text word DAVNO which has the dictionary meaning LONG. The routine functions only if DAVNO is concorded with a word functioning as predicate. In this case, DAVNO has its meaning modified, and the predicate is made perfective.

Before any other operations, the word order of the sentence is modified in those cases where DAVNO is concorded with a predicate to its left. If the predicate is a short form adjective or participle, the dictionary meaning LONG is retained and the predicate is made perfective by inserting HAVE/HAS (depending upon number) before and BEEN after DAVNO. The fourth verbal form (past participle) is used.

Example: DAVNO IZVESTNO, CTO VYDELENIE CISTYX KUL6TUR
SV4ZANO S BOL6WIMI TRUDNOST4MI.

IT HAS LONG BEEN KNOWN THAT THE ISOLATION OF
PURE CULTURES IS CONNECTED WITH LARGE DIFFICUL-
TIES.

If the predicace is verbal, it is always made perfective as described above. (If the verb is past tense, HAD is used instead of HAVE/HAS). The translation of DAVNO is rendered A COMPARATIVELY LONG TIME if it is preceded by SRAVNITEL6NO (which is eliminated from translation).

Example: 3TA PROBLEMA IZUCAETS4 SRAVNITEL6NO DAVNO.

THIS PROBLEM HAS FOR A COMPARATIVELY LONG TIME
BEEN STUDIED.

If the predicate is a past tense reflexive verb, DAVNO is
rendered LONG AGO.

Example: 3TO DAVNO 4VL4LOS6 USTANOVLENNYM FAKTOM.

THIS WAS LONG AGO AN ESTABLISHED FACT.

OBR

This routine is initiated by the text word OBRATNO, which
carries the dictionary meaning BACK (TG1 = INVERSELY). The
meaning INVERSELY is used if OBRATNO is directly followed by a
form of the adjective PROPORQIONAL6NY1.

Example: POSLEDN44 MEN4ETS4 OBRATNO PROPORQIONAL6NO
CETVERTOL STEPENI DIAMETRA.

THE LATTER VARIES INVERSELY FROM THE MACHINE.

The meaning BACK is used if (1) OBRATNO is directly
followed by either of the prepositions V or NA governing the
accusative, or if (2) OBRATNO is preceded by the conjunction
I which is preceded by the accusative object of either of the
prepositions V or NA, or if (3) OBRATNO is directly followed
by the preposition OT and within five words to the right there
follows the preposition K, or if (4) within the sub-sentence
unit to the left of OBRATNO a word is encountered beginning
with the clusters UXOD, PERE, or STEKA.

Example: ONI LETELI OBRATNO NA ZAPAD.

THEY FLEW BACK WEST.

or: PROISXODIT MIGRAQI4 MNOGIX VIDOV PTIQ, LET45IX
S MEST GNEZDOVYX NA ZIMOVKU I OBRATNO.

(THERE) OCCURS A MIGRATION OF MANY TYPES OF
BIRDS, FLYING FROM THE NESTING GROUNDS TO
WINTERING AREAS AND BACK.

or: REZUL6TAT DELSTVI4 PEREDAETS4 OBRATNO OT RUL4 K
KOMPASU.

THE RESULT OF THE ACTION IS TRANSFERRED BACK
FROM THE WHEEL TO THE COMPASS.

or: NA KRYWKE KONDENSIROVALAS6 ISPAR4H5A4S4 JIDKOST6
I STEKALA OBRATNO V BAK.

ON THE CEILING CONDENSED THE EVAPORATED LIQUID
AND FLOWED BACK INTO THE TANK.

The meaning VICE-VERSA is used if OBRATNO is immediately
preceded by the conjunction I but it is not preceded by the
accusative object of V or NA.

Example: 3TO PEREME5ENIE FRONTA PRIVODIT K IZMENENIH
XAKAKTERA OTVETOV NEIRONOV S TONICESKOGO NA
FAZICESKII I OBRATNO.

THIS SHIFT IN THE FRONT LEADS TO CHANGE IN THE
CHARACTER OF THE ANSWERS OF THE NEURONS FROM
"TONIC" TO "PHASAL" AND VICE-VERSA.

VES6MAR

This routine is initiated by the text words VES6MA (diction-
ary meaning AGAIN). When VES6MA is followed by a verbal, its
meaning is given as COMPLETELY.

Example: 3TO VES6MA OSLOJN4ET PRAKTIKI IMMUNIZAQII.

THIS COMPLETELY COMPLICATES THE PRACTICE OF
IMMUNIZATION.

When VES6MA is followed by a long form adjective or
participle, its meaning is VERY.

Example: 3TO MOJET BYT6 VES6MA OGRANICENNYM.

THIS CAN BE VERY LIMITED.

When VNOV6 is followed by a verbal, its meaning is made
AGAIN.

Example: POSLE CEGO VNOV6 PEREVIVALI NA VYWEUPOM4NUTYE
SREDY.

AFTER WHICH THEY AGAIN REWOVE ON THE AFORE-
MENTIONED MEDIUM.

When VNOV6 is followed by a long form adjective and
participle, its meaning is NEWLY.

Example: VNOV6 OTKRYVAEMYE ANTIBIOTIKI OKAZYVAHTS4 UJE
OPISANY.

NEWLY DISCOVERED ANTIBIOTICS ARE ALREADY
DESCRIBED.

c. PROPER NAMES. An important development effected during
this optimization phase has been the recognition and automatic
translation of proper names and initials. This is accomplished
by two routines: NAMER in Structural Pass 0 and NAMCHR in
Structural Pass 4.

NAMER is designed to establish "Name Units" of variable lengths in non-homogenous material. The initiating characteristic in the input material is initial capitalization. The environment is then checked for possible initials, given names and/or patronymics in the standard dictionary and/or titles of professions. This material can occur either in a normal [title + given name(s) + last name] sequence or in the reverse bibliographic order: [last name + given name(s)]. The difficulty is compounded by the variability in material, by the fact that last name and initials may either be not-found words or words with other parts of speech assigned in MDL (Main Dictionary Look-up), and by the fact that the sentence often contains only one capitalized last name without any initials or titles to confirm its function.

Many names are encountered for the first time in each new text and so must be processed automatically without inclusion in the standard dictionary. The "name units" must be identified and structurally linked for correct synthesis of the English text. This includes the correct placement of prepositions and articles (though these are seldom used). In addition, in order to assign the proper case, gender, and number to the "name units," the endings and environments of the names must be examined.

Thus, all the ambiguities created by dictionary status are overcome before structural analysis of such critical areas as noun government and subject-predicate relationships occurs.

NAMECHR further refines the treatment of those names which were not taken from the standard dictionary (the not-found words). In order to use the nominative-singular Russian as the normative English name, the bit structure after structural analysis is checked and the endings accordingly altered. Following this, transliteration of the name is carried out on the new stem to ensure correct assimilation of the name into English.

d. NUMERALS. Provisions for the structural analysis of literal and written cardinal and ordinal numerals are made in all five structural passes. The position of each routine is determined by the need in succeeding passes for the codes and relationships it establishes.

The two routines in Structural Pass 0, SPOR5 and SPOR6, set the preliminary codes necessary for the proper recognition of basic grammatical relationships in Structural Pass 1. The first routine, SPOR5, deals with the irregular government of the written cardinal numerals DVA, DVE, TRI, CETYRE, OBA, and OBE and of all literal cardinal numerals ending in 2, 3, or 4. It sets up the appropriate case, gender, and number codes for recognition in later passes of the concord of these numerals with following nouns. It also provides for the plural translation of the noun to be modified.

The second routine, SPOR6, performs an even more basic function: it examines not-found words in order to isolate those which are literal ordinal numerals (3-GO, 15-E, etc.).

The isolated numerals are then assigned the case, gender, and number codes appropriate for their endings and for the noun which follows them. Further provisions are made to translate the ordinal numerals with the appropriate English equivalent (3RD, 15TH, etc.).

In Structural Pass 1 the underlying structural relationships are established. The NUMBER10 routine has separate branches for literal cardinal numerals ending in %, for literal cardinal numerals ending in 1 (except 11), and for all other cardinal numerals. Ordinal numerals are treated as adjectives for the analysis of potential concord. Literal cardinal numerals ending in % are placed in an adnominal genitive relationship if the noun following is in the genitive. Otherwise they proceed through the routine for all other cardinal numerals.

The OBAR in Structural Pass 2 completes the handling of the irregular government of OBA, OBE, DVA, DVE, TRI, and CETYRE which was begun in Structural Pass 0, SPOR5. The cross reference codes for this relationship are inserted in this routine.

In Structural Pass 3, cardinal numerals are included in the list of subject scans. A literal cardinal numeral is accepted as a subject if it is neither governed nor an appositive and if the predicate is not a short form neuter singular adjective. A written cardinal numeral must be in the nominative case or govern or modify a noun or govern the preposition

IZ to be accepted as a subject. Special provisions are made to ensure that the numeral subject agrees in number with the English predicate.

ORDNUM in Structural Pass 4 was written to correct the possible concord of a cardinal numeral with a noun phrase modified by an ordinal numeral and to provide for the insertion of the particle "OF" before an ordinal numeral which does not modify a noun.

The execution of these structural routines, linking each cardinal or ordinal numeral, whether written or literal, with a noun or verb with which it is syntactically related, often determines the corresponding English translation.

e. ADNOMINAL GENITIVES. During the systematic analysis of SYSTRAN output, the shift towards sophisticated problems became apparent. Disregarding dictionary problems, which are simultaneously corrected during the output analysis procedure, the factors which influence translation are those complex situations which come to the surface only when a system reaches a high level of automatic parsing capability.

The following example is an illustration of the higher level problems. (Due to the length of the sentence, only the section relevant to the problem will be given.)

Example: . . . NA VOZNIKOVENIE POSTO4NN01 SOSTAVL4H5E1
SKOROSTI VETRA V NIJNIX SLO4X . . .

The SYSTRAN translation is:

. . . ON THE EMERGENCE DIRECT COMPONENT WIND VELOCITY IN
THE LOWER IONOSPHERE LAYERS . . .

It is obvious from reading this translation that the adnominal genitive relationship between SOSTAVL4H5E1 and SKOROSTI was not at first recognized. Upon analyzing the printout of the sentence analysis area in the computer high speed core memory, at the time when the translation took place, it became clear that both POSTO4NN01 and SOSTAVL4H5E1 were recognized as modifiers of SKOROSTI, due to the fact that they have common cases and genders. Their part of speech reflects that they are noun-adjective ambiguities. If such noun-adjective words can modify a noun, according to the present rules such syntactic connections should have preference.

On the basis of the above example, the following additional rule was added: If a noun-adjective is preceded by a modifier and followed by a noun, an adnominal genitive relationship should be preferred (if codes allow it) rather than a modifier relationship. After the incorporation of this rule, the above sentence portion was translated as follows:

. . . ON THE EMERGENCE OF THE DIRECT COMPONENT OF WIND
VELOCITY IN THE LOWER IONOSPHERE LAYERS . . .

Many similar higher level problems have been encountered. In all cases the modularity of SYSTRAN made the easy inclusion of the improvement possible.

f. VERBS + OBJECTS. Instances have been encountered demonstrating that some Russian verbs may function transitively and that they may also, without any change in their lexical meanings, require a specific preposition which expresses the verb-object relationship by assuming the role of the governor. As an example, the verb SUDIT6 is considered here and two sentences are given to illustrate the two grammatical phenomena.

First, SUDIT6 is used as a transitive verb requiring its object in the accusative case.

ONA STAVIT SEBE QEL6H IZLOJIT6 VPECATLENI4 OCEVIDQA,
SKRYVAVWEGO SVOE IM4 POD NAZVANNYM PSEVDONIMOM OT RABOTY
TOGO "BAKTERIOLOGICESKOGO OTR4DA C\$KVANTUNSKOI ARMII",
DE4TELEI KOTOROGO SUDILI NA XABAROVSKOM PROESSE.

Second, SUDIT6 is used with the preposition OB which governs the object in the locative case.

TOL6KO TAKOE SRAVNITEL6NOE ISSLEDOVANIE POZVOLIT SUDIT6
OB ISTINNOM ZNACENII TOGO ILI INOGO KOMPONENTA KLETKI V
VIRULENTNOSTI KUL6TURY.

While in the first example the direct-object requirement is treated in the normal manner during the structural analysis, in the second example this requirement must be suppressed. An instruction must be implemented which shuts off the accusative requirement once the preposition O or OB is encountered, thereby inhibiting the recognition of ISSLEDOVANIE as the direct object of the verb. The subsequent subject routine can then

correctly recognize it as the subject of the sentence.

This example leads to the hypothesis that this function of the prepositional phrase initiated by O, OB may be related to the fact that a proper English translation of the preposition is simply the omission of its meaning (SYSTRAN's preposition-translation code 99). When the meaning of a preposition is omitted, the English usage corresponds to an English direct object, the equivalent of a Russian accusative. Thus in each case when a verb governs a preposition and indicates an omitted translation for that word, the accusative requirement should be suppressed.

g. SEQUENCE OF TENSES. During the output analysis it was revealed that sentences containing certain adverbial expressions require special considerations with respect to the translation of tense of the finite verb used in the particular sentence. Two such expressions were noted and classified as a type of indicator signalling that the present tense in the Russian narrative might be rendered by the past tense into English. The words were POSLE and ZATEM, translated AFTER and THEN, respectively.

The English AFTER may be followed by the present tense under limited conditions only. It has been observed that in Russian narrative in which the author deals with the description of an experiment or a scientific event which took place in the past, the finite verb frequently occurs in the present

tense after prepositional phrases initiated by POSLE and ZATEM. In English, however, in order to prevent any interruption in the continuity of action with respect to time, these present-tense finite verbs have been regularly rendered in the simple past tense. The tense-conversion step can be accomplished before the English Synthesis program is executed. Examples follow:

CEREZ NEKOTOROE VREM4 POSLE OTDELENI4 AVTOMATICESKOL
STANQII OT RAZGONNOGO RAKETNOGO BLOKA VKLHCAETS4 SISTEMA
ORIENTAQII.

AT A CERTAIN TIME AFTER THE SEPARATION OF THE AUTOMATIC
STATION FROM THE ACCELERATING ROCKET BLOCK THE ORIENTA-
TION SYSTEM WAS SWITCHED ON.

POSLE POLUCENI4 SVETLOGO RASSYPCATOGO WLAKA VVOD4T
ALHMINI1 ILI FERROALHMINI1.

AFTER OBTAINING A WHITE FALLING SLAG, ALUMINUM OR FERRO-
ALUMINUM WAS ADDED.

The adverb ZATEM appears to create a similar situation, especially when the present-tense finite verb is in the third person plural.

ZATEM VYCISL4HT SODERJANIE KOBAL6TA I JELEZA V
METALLICESKOM NIKELE.

THEN THE CONTENTS OF COBALT AND IRON WERE CALCULATED.

ZATEM PRIBAVL4HT =N ML VODY, =N G MOCEVINY I NAGREVAHT =N MIN.

THEN IT WAS TREATED WITH =N ML OF WATER AND =N G OF UREA, AND HEATED FOR =N MIN.

In addition, other words which seem to influence the choice of tense have been considered. The expression SLEDUET/IT FOLLOWS indicates that the future tense in the Russian sentence should be translated by the present in English.

IZ /=N/, KAK I IZ /=N/, SLEDUET, CTO 3TO USLOVIE BUDET VYPOLN4T6S4 TUT V DOSTATOCNO BOL6WOM RAZMERE TELA.

FROM =N AS FROM =N, IT FOLLOWS THAT THIS CONDITION IS FULFILLED FOR A BODY OF SUFFICIENT SIZE.

Also, the connector ESLI when followed in Russian by the future-tense verb form (imperfective or perfective aspect) should be added to the list of words to be considered for rendering the present-tense verb in English.

ESLI VY4SNITS4, CTO EGO POKUPAHT DL4 LECENI4 PONOSA, . . .

IF IT TURNS OUT, THAT THEY BUY IT FOR THERAPY OF DIARRHEA, . . .

h. ABBREVIATIONS. In order to attain high efficiency in translating abbreviations by automated means, a more detailed processing was devised and made effective. Prior to these measures each abbreviation was assigned a code by which it was identified by POS and BPQ in the dictionary and processed during structural analysis as a common noun. No provisions,

however, were made for further considerations on the syntactic level. This approach was too generalized and, in terms of the more sophisticated development of the System in other areas of the linguistic aspects of the Russian language, no longer adequate. Improvements were therefore instituted.

The main effort was to review the existing file of abbreviations and to recode each entry along carefully designed lines and, at the same time, to develop a methodology that could be expanded and utilized either for coding the not-found abbreviations revealed during each translation run or for systematically coding a dictionary of abbreviations. (The latter would require the expansion of input conventions to allow for better differentiation of entries through the identification of italics, variable capitalization, etc.)

As a result of a thorough review of the existing file, it was found that the standard POS/BPQ of LC/13 should be retained only for those abbreviations which perform a noun function in the sentence. In this way, structural modifications can be made to accommodate this large group more specifically. What did emerge from the investigation was that in addition to the noun function, abbreviations, depending on interrelations of words in a sentence, also function as adjectives, adjectival nouns, pronoun-adjectives, proper nouns, titles of professions, superlative adjectives, adverbs, prepositions, participles, numerals, and verbals. Steps were taken to recode the file according to these diverse functions.

During the analysis it also became clear that certain abbreviations needed to be considered as homographs and should therefore bear codes indicating their candidacy for homograph analysis. It is anticipated that this approach will prove itself rewarding and will increase in significance in the future as more abbreviations are encountered. To illustrate the importance of correctly coding items which, due to their versatility of function, may have an impact on their environment as far as the grammatical structure is concerned, the Russian SM is here considered. It has been encountered with two of the following meanings: (1) CENTIMETER or CM., and (2) the imperative form of "see," i.e., SEE, q.v., vide, or cf. When used as a unit of measure, this abbreviation is preceded by a numeral to which it is syntactically tied. When it functions as a verbal, however, it is desirable that the System recognize it as such, since the structural analysis should reveal that SM. is a predicate, or governs a direct object, or requires certain prepositions. The sentence structural analysis will be greatly improved as a result of this, yielding a better translation.

A list is given below presenting the various noun functions of abbreviations as they were encountered during machine translation output analysis. The POS and BPQ codes assigned to these entries in SYSTRAN dictionaries are at the head of each group of Russian abbreviations and their English equivalents.

Pure Adjective

POS/BPQ = 20/20

ABS.
AMER. }
AMERIK.
AMERIKAN }

ANGL.
AVTODOR.
BYV.
C.D.A.
FIZIC.
FOTOGR.
GEKSAG.
GQK
HJ.
IK.
INFORM.
KRIT.
KUB.
LAT.
MEXAN. }
MEXANIC. }

MN.
MOL.
NAST.
NELTR.
OTN.
OTRIQAT.
PER.
PEREM.
PRIKL.
QILINDRIC.
RUS.
SPEQ.
TEXNIC.

UCEBN.
UD.
U.F.
USL.
UZ
VOST.
X.C.
XIMIC.

ABS.
C\$AMER.

C\$ENG.
HWY.
FORM.
CC\$A.P.
PHYS.
PHOTO.
HEXAG.
CC\$FCC
C\$S.
CC\$I.R.
INFO.
CRIT.
CU.
C\$LAT.
MECH.

PL.
MOL.
THIS
NEUT.
REL.
NEG.
TRANS.
VAR.
APPL.
CYLIND.
C\$RUSS.
SPEC.
TECH.

EDCNL.
SPEC.
CC\$U.V.
CONV.
US
CC\$E.
CC\$C.P.
CHEM.

3KSPERIM.
 3KSPONENQ.
 3L.-GRAFIC.
 3L.-MIKROSKOPIC
 3L.-OPTIC.
 3LEKTRIC.
 3LEKTROSTATIC.
 3NERG.
 3NERGETIC.
 4DERN.

EXPERIM.
 EXPON.
 EL.-DIFF.
 EL.-MICRO.
 EL.-OPT.
 ELEC.
 ELECTROSTAT.
 C\$E.
 ENERG.
 NUCL.

Adjectival Noun

POS/BPQ = 24/21

If the noun and the adjective meanings differ in English, the adjective precedes the noun.

ARIFM.
 AT.
 ATM.
 C.M.
 ED.
 FILOS.
 FIZ.
 FR.
 GL.
 IMP.
 ISPYT.
 J.-D.
 J.D.
 J./D. }

KN.
 KV.
 LIT.
 MATEM. }
 MAT.
 MATEMAT.)
 MED.
 MEX.
 NC
 REG.
 SOKR.
 TEXN.
 UKV
 VC
 XIM.
 XROMATOGR.

ARITH.
 AT.
 AT.
 F/M
 SING./C\$U.
 PHILOS.
 PHYS.
 C\$FR./FR.
 MAIN/C\$CH.
 IMP.
 EXP.
 CC\$R.R.

BOOK
 SQ.
 LIT.

MATH.

MED.
 MECH.
 L.F.
 REG.
 ABBREV.
 TECH.
 CC\$USW
 CC\$H.F.
 CHEM.
 CHROMATOG.

Proper Names--Inanimate

POS/BPQ = 14/11

ARMSSR	C\$ARM CC\$SSR
BSSR	CC\$BSSR
FRG	CC\$FRG
GDR	CC\$GDR
GSSR	CC\$GSSR
KAZSSR	C\$KAZAKH CC\$SSR
KNDR	CC\$KDPR
KNR	CC\$CPR
KSSR	CC\$KAZAKH CC\$SSR
LITSSR	C\$LIT CC\$SSR
M.-L.	C\$M.-C\$L.
R.N.R.	CC\$R.P.R.
RSFSR	CC\$RSFSR
SSSR	CC\$USSR
SWA	CC\$USA
USSR	C\$UKR CC\$SSR
3SSR	CC\$ESSR
3STSSR	C\$EST CC\$SSR

Titles of Professions

POS/BPQ = 18/12

CL.-KOR.	C\$CORR.-C\$MEM.
INJ.	ENG.
INJ.-MEX.	MECH. ENG.
INJ.-TEXNOL.	TECH. ENG.
KAND.	MASTER
PROF.	PROF.

Superlative Adjective

POS/BPQ = 20/28

VYSW.	HIGH (codes allow synthesis in ESYN)
-------	---

Adverb--Never Predicate

POS/BPQ = 30/3F

D.S.P.	UNTIL NOW
D.B.	PROB.
GL.OBR.	MAINLY
M.B.	PERHAPS
PRIBL.	APPROX.

Preposition Requiring the Genitive Case

POS/BPQ = 50/BA

IM.	NAMED AFTER
VM.	INSTEAD OF

Preposition 'OKOLO'

POS/BPQ = 50/A8

OK.	ABOUT
-----	-------

Pronoun-Adjective

POS/BPQ = 28/22

K.-N.	SOME
-------	------

Past-Passive Participle

POS/BPQ = 80/83

GRANEQENTRIR.	FACE-CENTERED (and codes to synthesize it in ESYN)
---------------	--

Cardinal Numeral

POS/BPQ = AO/18

MLRD.
MLN.
TYS.

BILL.
MILL.
THOUS.

Imperative Form of the Verb

POS/BPQ = 04/48

SM.

SEE (and codes to synthesize it in ESYN)
(First processed as HM).

SR.

COMP.

In addition to the items listed above, all SYSTRAN entries coded as POS/BPQ = 1C/13 were updated and now bear codes indicating gender, number, case, animation, preposition translation, and any other applicable information. The purpose of all these codes is to ensure a more correct structural analysis and subsequently a higher-quality translation.

i. ARTICLES. The article routine has been improved, whenever logically feasible, to consider the following possibilities in translating from the Russian language to English:

1. No article should be inserted in the translation.
2. The definite article should be inserted.
3. The indefinite article should be inserted.

Several considerations must be utilized in the process of developing a workable tool which most dependably inserts the English article correctly. In scientific literature, the definite article THE is used more often than the indefinite article A or AN. With nouns standing alone, THE is used as much as 95 percent of the time while with others, use of THE may reach a full 100 percent: in one instance, ABSORPTION received THE 225 times, but AN only 12 times; ABSENCE received THE 135 times and AN not at all. Obviously, THE became the primary target of investigation.

Nouns standing alone may, if they belong to particular conceptual categories, require the determination normally expressed in English by THE. This article, although hardly subject to any collocational restrictions (articles are found in almost any lexical environment in the language that the grammar permits), can be shown as functioning "structurally"--outside any grammatical reference--in English. This is substantiated by observing article occurrence in a large corpus of running

text. This fact alone, provided the field is investigated exhaustively, or nearly so, eliminates the normally expected search for some manifest relationship to the referential and situational meaning of the words concerned. The range is therefore narrowed to the ultimate or, at the most, immediate constituent. This is primarily possible because the literary style of technical and scientific publications tends to make use of sequence-type sentences rather than of situation-type ones.

The information below is the result of a preliminary analysis of nouns and their determiners. The example shows nouns that prefer, to a certain extent, the definite article.

Name	No. of Occurrence	THE %	A/AN %
ABILITY	48	100	-
ABSCISSA(S)	24	100	-
ABSENCE	135	100	-
ACCUMULATION	42	93	7
ABSORPTION	267	95	5
VALUE(S)	102	95	5

Another way of considering the determiner is to look for the presence of a formula or a number which, together with the noun that follows it, forms an identifiable unit. The definite article is placed before such a unit as the following examples illustrate (in the transcription formulas are rendered by "=X"

and numbers by "=N") :

THE =N =X LEVEL
THE =N =X REGION
THE =N =X WAVELENGTH
THE =N AND =N ENZYMES
THE =N AND =N LEVELS
THE =N AND =N STATES
THE =N AND =N SUBLEVELS
THE =N AND =N SOLUTION AMPLIFIERS
THE =N CENTURY
THE =N COEFFICIENTS
THE =N DERIVATIVES
THE =N DIRECTION
THE =N DONORS
THE =N ELECTRODE
THE =N EQUATIONS
THE =N EXPERIMENTS
THE =N LEVEL
THE =N MASS SPECTROMETER
THE =N METHOD
THE =N NETWORK
THE =N NEURONES
THE =N OXAZOLIDONE
THE =N PENICILLIUM STRAINS
THE =N PENICILLIUM CULTURES
THE =N PHAGE
THE =N PHAGOLYZATE
THE =N PHOTOGRAPHS
THE =N PHOYPHAGE
THE =N PLANE(S)
THE =N POLYGONS
THE =N POLYPHAGE
THE =N POSITION
THE =N PYRAZOLINE
THE =N RADIATION
THE =N RANGE
THE =N REFLECTION
THE =N REVOLUTION
THE =N RIBOFLAVIN
THE =N SECTION(S)
THE =N SIGNIFICANCE LEVEL
THE =N SPECTROMETER
THE =N STATE(S)
THE =N STERIOD DERIVATIVES
THE =N SUBLEVEL(S)
THE =N TRANSITION
THE =N TYPE
THE =N VALVE
THE =N YEAR

THE =N - DAY EXPERIMENT
 THE =N - NEURONES
 THE =N - PHENYL
 THE =N - POSITION
 THE =N - TYPE
 THE =N - UNIT
 THE =N - VECTOR
 THE =N - NND DAY
 THE =N - NND DEGREE
 THE =NST CENTURY
 THE =NTH CHANNEL
 THE =X AND =X ALLOYS
 THE =X AND =X AXES
 THE =X AND =X BRANCHES
 THE =X AND =X CONTENTS
 THE =X AND =X CHROMATOGRAPHS
 THE =X AND =X COMPONENTS
 THE =X AND =X CULTURES
 THE =X AND =X CELL(S)
 THE =X AND =X ENZYMES
 THE =X AND =X FORMS
 THE =X AND =X LEVELS
 THE =X AND =X MOLECULES
 THE =X AND =X MODES
 THE =X AND =X PARAMETERS
 THE =X AND =X RATIOS
 THE =X AND =X SYSTEMS
 THE =X AND =X SUBLEVELS
 THE =X AND =X TYPES
 THE =X AND =X VALUES
 THE =X ALLOY
 THE =X ARE + complement
 THE =X AXES
 THE =X AXIS
 THE =X BAND(S)
 THE =X CELL(S)
 THE =X COMPONENT
 THE =X COORDINATE(S)
 THE =X CULTURE
 THE =X CURVE
 THE =X CYCLE
 THE =X DERIVATIVE
 THE =X DEVICE
 THE =X DOUBLET
 THE =X ELECTRODE
 THE =X ENZYME(S)
 THE =X FORM
 THE =X GROUP
 THE =X ION(S)
 THE =X IS + participle or complement
 THE =X LEVEL(S)

THE =X LINE(S)
 THE =X MARTENSITE
 THE =X MOLECULE(S)
 THE =X PHAGE
 THE =X RADIATION
 THE =X RATIO
 THE =X REACTION
 THE =X REGION(S)
 THE =X REPRESENTATION
 THE =X SALT
 THE =X SPECTRUM
 THE =X SPHERE
 THE =X STATE
 THE =X STRUCTURE
 THE =X SUPERSTRUCTURE
 THE =X TRANSFORMATION
 THE =X TYPE
 THE =X VALUE(S)
 THE =X VECTOR(S)
 THE =X WAS + participle or complement
 THE =X WERE + participle or complement
 THE =X --BRANCH
 THE =X --COMPONENT
 THE =X --ELECTRON
 THE =X --IRON
 THE =X --METER
 THE =X --PARALLAXES
 THE =X --PHASE
 THE =X --PLANE
 THE =X --REGION
 THE =X --SPACE
 THE =X --TH + noun
 THE =X --TYPE
 THE =X --ZONE

Work on the definite article in English also consisted of examining a large body of texts for the occurrence of THE before words placed in complete sentences as modifiers of nouns. The aim was to establish groups of words grammatically used as adjectives which by themselves could adequately define the noun, necessitating the insertion of the definite article before it. To avoid an approach considered nugatory in any manner, the procedure was based on live and valid data which provided for

a circumspect observation of the "ambivalence" generally attributed to the English articles. The outcome yielded empirically valuable figures which were systematically compiled for the improvement of the article insertion program.

The search for modifiers influencing the choice of the definite article went, due to the availability of factual sentences taken from texts of various scientific disciplines, beyond the conventional selection of adjectives (NEXT, SAME, VERY, ONLY, PAST, ORIGINAL, etc.) by carefully looking at modifiers which were consistently preceded by the definite article and at modifiers which consistently prohibit the presence of the article (EACH, EVERY, ALL, etc.) The table below shows words which when modifying nouns are preceded by THE.

Name	No. of occurrence	THE %
ACHILES + N	12	100
ACOUSTIC + N	37	100
ADJACENT	22	100
ADJOINT	12	100
AFOREMENTIONED	11	100
AGGREGATED (+N)	3	100
ALLOWED (+N)	5	100
ALPHA (+N)	10	100
ANTIBIOTIC	22	100
ANTICIPATORY	10	100
APPROXIMATING	5	100
ARITHMETIC(AL)	12	100
ASCENDING	26	100
ASSOCIATED	5	100
AUDITORY	12	100
AUSTENTITE	19	100
AVAILABLE	15	100
BACTERIAL	39	100
BAROMETRIC	4	100

Name	No. of occurrence	THE %
BARATROPIC	4	100
BASE	41	100
BEST	81	100
BIOCHEMICAL	8	100
BLOOD	11	100
C\$BOOLEAN	20	100
BREAKING	20	100
C\$BROWNIAN	10	100
CALCULATED	85	100
CARBIDE	16	100
C\$CASPIAN	22	100
CAULOBACTOR	7	100
CELLULOLYTIC	11	100
COMPUTATIONAL	13	100
COMPUTED	16	100
CONTRALATERAL	12	100
CORONARY	11	100
CORRECTED	9	100
C\$DEBYE-C\$HUCKEL	9	100
DECARBOXYLASE	5	100
DECIMAL	9	100
DEFORMED	13	100
DEHYDROGENASE	11	100
DEPOSITED	37	100
DESCENDING	10	100
DESCRIBED	36	100
DESIRED	38	100
DETECTING	12	100
DETERMINED	12	100
ACTION + N	79	90.0
ACTIVE	87	89.0
ACTUAL	76	91.6
ARIAL	59	90.8
AIR	13	76.5
ALLOWABLE	11	90.9
AMINO	31	93.9
ANGULAR	56	94.9
ANOMALOUS	24	86.0
APPROPRIATE	17	85.0
ASYMPTOTIC	47	81.0
ATOMIC	12	92.7
AUTOMATIC	58	83.0
AVERAGE	154	91.0
AXIAL	31	96.9
BASAL	22	91.7
BASIC	139	94.0
BINARY	37	86.0
BIOLOGICAL	28	90.7

Name	No. of occurrence	THE %
BOUNDARY	78	96.8
CAPILLARY	14	87.5
CARBON	61	87.2
CATHODE	14	77.8
CELLULAR	19	95.0
CENTRAL	64	85.4
CHEMICAL	31	96.9
COMBINED	18	90.0
COORDINATE	48	95.9
CORRECT	24	92.3
COSMIC	27	93.1
CRITICAL	155	95.1
C\$DEBYE	17	94.5
DIELECTRIC	27	96.3

When certain nouns are used in collocation-type constructions with adjectives, it is possible to group them according to conceptual categories to which they belong and to determine whether or not certain adjectives warrant use of the definite article. One such category is "parts of a living body" as shown by the examples below:

THE CONTRALATERAL LEG
 THE CONTRALATERAL BICEPS
 THE CONTRALATERAL LIMB
 THE CONTRALATERAL SIDE

Since the adjective also qualifies to be categorized conceptually, groups can be developed permitting the application of rules for the insertion of the definite article. One such category is "proper name" as in THE CASPIAN GROUND-WATER FLOW which is directly related to the name THE CASPIAN SEA. When the construction PROPER NOUN + NOUN is preceded by a modifier the whole phrase is considered as one unit and is

determined by THE, as in THE CRITICAL JOSEPHSON CURRENT.

Additional examples are offered below:

THE BOOLEAN COMBINATION
FUNCTION
MATRIX
MATRICES
SUBMARINES

THE BROWNIAN MOTION
PARTICLES

THE DEBYE FORMULA
FREQUENCY
RADIUS
SPHERE

THE DEBYE-HUCKLE THEORY

In cases where neither the individual noun nor the adjective can decisively resolve the dichotomy, the search for the signal is directed to both members of the collocation to see which of the particular phrases should be considered. Preliminary checks have revealed that METHOD and PHASE, whenever modified by ACOUSTIC and ASCENDING, respectively, received the indefinite article but, whenever modified by DESCRIBED and CARBIDE, received the definite article. Similarly, the phrase MACHINE TRANSLATION almost never takes any article, but TRANSLATION alone may take either THE or A.

It was noted that certain expressions were brought forward as evidence or support of the subject matter discussed by the author. These words, for the most part modifiers, were enclosed within quotation marks and the quotation mark

indicating the beginning of the "quote" was preceded by the definite article THE. Below is given a list of expressions extracted from scientific texts in which each of the expressions was preceded by the definite article only.

"ALL OR NONE" PRINCIPLE
"ABBREVIATED" SYSTEM
"ACTION SPECTRUM" OF . . .
"ACTIVE" STATE
"ADULT" ONES
"AFTER EFFECT" IN . . .
"ANNEALING" CURVE
"AVERAGE" OBSERVER
"BACKGROUND"
"BEAT" TYPE
"BIOLOGICAL" CLOCK
"CELLS"
"CENTRAL" ION
"CENTRAL" PARTICLE
"CHARGING" OF . . .
"CLOCK"
"COEFFICIENT OF MAXIMUM OF VARIANCE"
"CONE" INSTABILITY
"CONSTANT" CORRECTION
"CONSTRUCTION"
"CONTRIBUTION" OF . . .
"CORRECT" REPLACEMENTS
"CRITICAL MASS"
"DARK" CULTURES
"DARK" COEFFICIENTS
"DARK" VALUES
"DERIVATIVES" OF . . .
"DIOXANE"
"DISCARDS"
"DISSOLUTION" OF . . .
"EFFECT" FUNCTION
"EFFECTIVE FETCH DIRECTION"
"ELECTRON" TEMPERATURE
"ELECTROSTATIC" COMPONENT
"ENZYMATIC" SIGNAL
"FILAMENTOUS" BACTERIA
"FIRST" CASE
"FORCE" OF . . .
"FORM =N"
"FORM =X"
"FROZEN-IN" SYSTEM
"GAS" OF THE SEGMENTS

"GENERALIZED HAMILTONIAN"
 "GULF STREAM =N"
 "HALF-MOON" ARRANGEMENT
 "HALF-MOON" CHROMATIN ARRANGEMENT
 "HORIZONTAL EDGE"
 "HUMP" IN THE CHARACTERISTICS
 "HYDROPHILICITY"
 "INCOHERANT" RADIOWAVE
 "INTACT" ZONE
 "INTERNAL CURRENT"
 "LABELED" BASES
 "LATE RECEPTOR POTENTIAL" =X
 "LATE" =X
 "LIGHT" CULTURES
 "LIGHTNING" MECHANISM
 "LOCKING" OF THE IONS
 "LONGWAVE" FORMS
 "LOSS CONE INSTABILITY"
 "MARGIN COEFFICIENT"
 "MARKUSSONOVSKII" FRACTION
 "MATURE" STATE
 "MAXIMUM SODIUM PERMEABILITY"
 "MEASURED" SLANT RANGES =X
 "MINIMUM" PRINCIPLE
 "MINIMUM-B" PRINCIPLE
 "MINIMUM-B" TYPE
 "MINIMUM-B" CONDITION
 "MODULAR DERIVATIVE" OF . . .
 "MONOMERIC" FORMS
 "NARROW" SITES
 "OPERATION CABOT"
 "OXYGEN" EFFECT
 "PEROXIDE" RADICALS
 "PHASIC" MONITOR UNITS
 "PLATEAU" REGION
 "POLAR" TIDE
 "PROHIBITED" ZONE
 "PROHIBITED" INTERVAL
 "RADIATION" PROFILE
 "RED BAND"
 "RED DROP"
 "RED" BAND
 "REMAINDER"
 "RESPONSE"
 "RETINA"
 "RIGHT HAND ROTATION"
 "RODINA- =X" TYPE
 "SATURATED" ONES
 "SATURATED" SYNTHETIC
 "SECOND" PHASE
 "SHEAR"

"SHOULDERS"
"SITE"
"SKELETOMUSCULAR" ONES
"SPIKES"
"STABILITY MARGIN"
"STANDARD LIST"
"START LINE"
"STRELA" DIGITAL COMPUTER"
"STRENGTH - DURATION" CURVE
"SUBLAYER"
"TENSOR OF THE HEART"
"TETANUS"
"TOTAL VARIATION"
"TRANSPARENCY INDEX"
"TRANSITIONAL" SPECTRA
"UNDAMAGED" SIDE
"VARIABLE COMPONENT"
"VELOCITY" SPREAD
"VERTICAL COSMIC PROBE
"WARM AND COLD PATCHES"
"WHOLE" CHARACTER
"WORKING" INTERVAL
"YES-NO" TYPE
"ZAZER'S" EXPERIMENTAL PLOTS

The given list establishes the fact that quotation marks are used abundantly in scientific texts. The palpability of the information warrants a special classification of the quoted expressions for the purpose of establishing the predictability of the occurrence of the definite article.

Codes have been established on the basis of the results of these investigations and incorporated into the System. It has also been possible to code verbs which, when functioning predicatively, define nouns in such a way that definite articles can be predicted for these nouns. Moreover, certain prepositions (e.g., DL4 and V) have been coded as indicators of definite articles for their objects.

SECTION IV

PROGRAMMING SPECIFICS IN IMPROVING THE EFFICIENCY OF SYSTRAN PROCESSING

1. SELECTION OF L. S. COMPOUNDS WITHIN TOPICAL GLOSSARIES

Original implementation of the Limited Semantics Compound Dictionary did not allow for special selection of compounds by topical glossary. As the number of L. S. expressions added to the SYSTRAN master dictionary grew, a provision was made to carry topical glossary indications; however, when the compound dictionary was created for use during translation processing, all expressions were considered as belonging to the general meaning category.

By including all expressions in the dictionary, regardless of topical glossary for a common principal word, the generated compound record became extremely large. Consequently, MDL (Main Dictionary Lookup program) was not able to attach the long compound information within the limits selected for efficient processing through the sort following MDL and prior to the final translation phase. The long compound information had to be retrieved by the LSLOOKUP program in the final translation phase. Additionally, selection of matching compounds with the text is more accurate in that the compounds are searched for a match in the priority order specified for the text being processed.

To create the L. S. Compound disk dictionary by topical

glossary, the CLS4 program was modified. This modification included the creation of an index file to the main dictionary file. An index record was created for each principal word and contained the topical glossary indicator, disk address of the corresponding compound record, and the length of the compound record. The disk address of this index record rather than the disk address of the compound record is passed to DUPDTLS (L. S. address updating program) for including the address in the master stem dictionary file. MDL was then modified to access this index record. The index record is scanned for requested topical glossary information, and, using the length specification in the index, the compound record is either attached to the text word or the disk address for later retrieval by LSLOOKUP.

The LSLOOKUP program was modified to analyze the compound records for a match with the text in requested topical glossary sequence. Three disk track areas are provided for retention of compound information, and, prior to accessing the disk file for compound information, these areas are examined for the presence of the desired compound already residing in memory, thus eliminating disk retrieval time.

Parallel to making the before mentioned program changes, the L. S. compound creation programs CLS1 and CLS2 were eliminated and their functions included in the translation processing programs LOADTXT and MDL since their contents were almost identical. By combining the instructions in common programs, any

change to the basic mechanics of processing can be made to just one program, eliminating updating and assembling of two separate programs. Additionally, this method assures accuracy in performing identical dictionary lookup of text words being translated and of L. S. Expression words for the creation of the compound dictionary.

2. LEXICAL ROUTINES: IMPROVED HANDLING OF MULTIPLE-MEANING WORDS

In that the selection of lexical routines to be activated for text words being translated is partly based on the beginning word characters of the text word, a standard binary search routine could not have been used. The lexical control program was therefore initially designed with a sequential table lookup for the routine desired. The number of lexical routines written at that time were few so this method was about as efficient as the binary search type.

As the number of lexical routines increased to several hundred, the sequential search became time consuming. Therefore, a special binary search routine was added to the lexical control program to determine whether or not the text word has already been translated by prior translation programs, in which case execution of a lexical routine for this text word can be eliminated.

SECTION V

SYSTRAN AND THE CONTROVERSY ABOUT THE FEASIBILITY OF MACHINE TRANSLATION

SYSTRAN is being developed at a time of considerable controversy about machine translation, or more specifically, about the expected quality of machine translation, in the United States as well as abroad. In the United States a series of conferences were held at the U. of Texas, devoted to the study of high quality translation. In the Soviet Union, the journal Naučno-Tekničeskaja Informacija devoted a whole article in its 6th issue in 1970 to the evaluation of the quality of machine translation.

The Texas conference investigated the problem from both philosophical and technical points of view; we concern ourselves with the latter in this technical report. There seemed to be general agreement that subject matter (knowledge) can be transferred with fidelity by automatic means.¹ It was also agreed that translators cannot keep up with rapid technological advances and changes in technical terms. A large scale machine translation system can be updated, every week if necessary, with the latest terminology thus assuring correct and unvarying transla-

¹Final Report, "Feasibility Study on Fully Automatic High Quality Translation", University of Texas at Austin, Contract F30602-70-C-0129, p. 7.

tion of technical terms and concepts.

SYSTRAN uses topical glossaries to insure proper translation of technical terms which vary in meaning from one subject area to another. Not even one third of SYSTRAN's topical glossarization capability is used at this time. At present, meanings are attached to SYSTRAN dictionaries in only 10 subject fields, although the present System could handle 36 subject field areas on individual lexical entries as well as on compound levels. This means that SYSTRAN is presently equipped to translate texts in ten different specialized subject areas, and is potentially ready to translate in thirty-six areas.

The Texas report states that "programming of many linguistic algorithms remains a slow and difficult task."² It should be noted in this respect that this statement can be maintained only if a suitable language to communicate with the computer does not exist. The SYSTRAN MACRO language developed by LATSEC, INC. overcomes this hurdle and makes the programming of linguistic algorithms a quick and simple task which can be carried out by linguistic personnel themselves after a brief training period. The MACRO instructions are converted into direct assembly language, allowing communication with the computer in the language closest to the computer, thereby making full use of the logical capabilities and the available high speed core of the computer.

²Op.cit., p. 14.

There are two further prime considerations in discussing machine translation. One is time: "The need for virtually immediate translation is one of the major reasons for concern with machine translation."³ In other words, given a choice between overnight machine translation and a human translation in a fortnight, today's scientist, cognizant of the problem of obsolescence in scientific and technical writing, will tend to choose the perhaps less elegant, but certainly more immediate, automatic translation.

The other consideration is quality. (Nida and Bar-Hillel have both remarked that the level of quality desired is relative to the user.) For example, a physicist may be satisfied with a faithful, if labored, transmission of concepts, while a stylistician will require the idiom of the target language.

SYSTRAN's listing of L.S. compounds and idioms, in addition to its numerous specialized lexical routines, allow the linguist to greatly minimize stilted, too-literal translations by making special provisions to attain translation equivalency for idiosyncratic units or constructions as well as for individual lexical items. An example of this is the handling of the Russian MOJNO construction. Consider the following Russian sentence:

NEKOTORYE FAKTY MOJNO UJE SCITAT6 USTANOVLENNYMI.

There is in this sentence, no overt surface subject, since

³Op.cit., p. 6.

MOJNO, an impersonal expression, functions as subject and predicate. A literal, word-by-word translation could produce the following:

SOME FACTS (IT) IS POSSIBLE (or CAN) ALREADY TO CONSIDER ESTABLISHED.

Rearranging that translation would yield the following slightly more readable sentence:

IT IS POSSIBLE TO CONSIDER SOME FACTS ALREADY ESTABLISHED.

SYSTRAN, rather than supplying an impersonal subject such as ONE or IT, chooses the accusative object noun to serve as surface subject, yielding the following translation:

SOME FACTS CAN BE CONSIDERED ALREADY ESTABLISHED.

The article in the Soviet publication, Naučno-Tekhnicheskaja Informacija (NTI), analyzes possible criteria which may be developed to evaluate human and machine translation. It stresses the fact that no real attempt has been made to objectively evaluate automatic and human translation. It also stresses the fact that human translation may demonstrate a good many inadequacies which usually do not occur in machine translation. The smoothness of the human translation in comparison to the roughness of machine translation usually influences any evaluation adversely against the latter. The article elaborates upon methods of evaluating translations and sets up various evaluation

scales. The article also refers to the English-Russian machine translation carried out in the Soviet Union.

Segments of this article which have been translated by SYSTRAN are given below. The translation reflects present SYSTRAN quality resulting from a careful update of grammar codes and English synthesis codes. The hundreds and hundreds of highly sensitive algorithms incorporated into SYSTRAN can carry out their tasks only if the entries of the supporting dictionaries are correctly supplied with all the necessary grammar codes. A systematic improvement on the grammar code level often requires analysis of the occurrence of certain lexical items in context, since the human mind is incapable of determining a priori all the possible syntactic and semantic relationships in which a given entry may function. The translation also illustrates where further refinements are necessary. For example, in the sentence--

3TA PROQEDURA NE DAET, ODNAKO, OTVETA NA VOPROS, KAK
VLI4ET TOT ILI INOI TIP OWIBKI PEREVODA NA OB5UH OQENKU
KACESTVA.

--the Russian preposition NA may be governed by the noun PEREVOD as well as by the verb VLI4T6. Since PEREVOD immediately preceeds NA, SYSTRAN recognized it as the governor of the preposition, translating the preposition as INTO. In this case, however, the requirement of the verb overrides that of the noun, and the translation of NA should have been suppressed.

SYSTRAN's translation is presented in the following reductions of actual printouts. The output reflects a 1403 printout

rather than a 360 printout which does not possess a colon, semicolon or quotation marks. Consequently where such punctuation should have occurred, blanks occur in the translation. Following the English translation, the Russian original, which served as direct input to SYSTRAN, is given.

The last translated paragraph of the Soviet article states that, for practical translation purposes, a full dictionary is indispensable, implying that this need be augmented only by simple algorithms. We cannot agree with this statement. Although a full dictionary supplied with correct grammar codes is of great importance, simple algorithms cannot carry out the translation task. SYSTRAN's algorithms analyze each sentence from every possible syntactic point of view and, in addition, utilize a certain amount of semantic information. Previous decisions are reconsidered repeatedly as a coding scheme of about 100 bytes (800 bits) is built up after each word to express the necessary syntactic ties between the members of each sentence and subsentence. The hierarchical analysis carried out by the algorithms, which use sequences of subroutines and nested subroutines to carry out high level logical decisions, can be called anything but simple.

SYSTRAN's most powerful feature, however, is the open-endedness of these sophisticated algorithms which makes them amenable to further refinement. These refinements can be made continually at either the syntactic or semantic level. It should be kept in mind, however, that syntax and semantics

are so closely interrelated that the construction of any adequate grammar will require that as much care be devoted to the development of the semantic component as is given the syntactic component. SYSTRAN's present stage of sophistication makes it abundantly clear that semantics can and must be accounted for in an automatic recognition grammar.

SUBJECT CODE 3.

THE AUTOMATION OF THE TRANSLATION OF TEXTS.

THE EVALUATION OF THE QUALITY OF MACHINE
TRANSLATION.

V. A. Moskovich, S. M. Vishnyakova.

The purpose of the work is the definition of the acceptable and permissible quality of machine translation provided both to postediting and to direct distribution without postediting. Such a study must be conducted for the clarification of the possibility of the practical use of experimental systems of machine translation (MT) and for the formulation of requirements for the quality of the translation which should be ensured by the improved algorithms of MT.

The question of the permissible quality of machine translation is today one of the central questions of the theory and practice of MT, whose solution will in many respects contribute the substantiated prediction of

further works in this field. ...

The results of these works in the evaluation of the quality of translation do not make it possible to give a clear answer to the question placed in these works: how satisfactory is machine translation being obtained today? On the basis of these works it is possible to make only that conclusion, that the texts machine translated, according to all characteristics are inferior to the texts translated by man.

In connection with this conclusion let us note our difference of opinion with the being practiced in abovementioned and similar to them works procedure of presentation in the same experiments of the texts translated by people, and the texts machine translated. The fact is that in the absence of the preliminary attribution of translation to a certain qualitative class it is difficult to judge, what the text of the translation sets forth. Up to now a basis for the comparison of the qualitative characteristics of ordinary and machine translation is absent. It is unclear how is differed the machine translation from the ordinary. human translation can possess a mass of deficiencies which can in principle be absent from machine translation (for example the hypertrophy of the

features of the individual style of a human translator in the translated text, etc.). On the other hand, the "smoothness" of the grammatical structure of human translation (even with its incomprehensibility) can influence the subjective evaluation of the test subjects towards overstatement. On the contrary, the roughness of grammatical structure, peculiar to machine translation, but unusual to man, can sharply lower the evaluation of translation for his clarity.

In this way, none of the previous works in the evaluation of the quality of translation can be recognized satisfactory in every respect. When developing the procedure of the evaluation of the quality of the translations of patent texts it was necessary to anew develop all procedures of experiments, taking into account all that is rational, which was proposed in the works of our predecessors.

The principles placed as the basis for the procedure developed by us, are reduced to the following: 1) the clear differentiation of three independent parameters of the measurement of the quality of the translation: the clarity of translation, correctness of the communication of the sense of original in translation and acceptability of translation as

a whole; 2) the carrying out of the experiments with texts which are characterized only by some single type of distortions; 3) the absence in the experiments of human translations; 4) the neutralization of the factors, capable of distorting the real evaluation of translation; 5) the utilization of the scales with the optimum scale of the gradation of indicators; 6) the development of the positive establishment of the test subjects for machine translation before the beginning of an experiment.

The parameters of the measurement of the quality of translation chosen by us are differed by relative independence and the important aspects of the translation characterize: its clarity, the faithfulness to the original and its acceptability. The clarity of translation is basic from the evaluated parameters of the quality of translation. To measure it is possible according to the scales of the subjective evaluations of the "clearness of the sense" of translation.

The second parameter of the evaluation of the quality of translation - its faithfulness to the original - is in principle independent from the first parameter. It is also measured on the basis of the subjective evaluations of the test subjects.

Finally, the third parameter - the acceptability of translation - is the derived from the first two. In its measurement being accomplished also according to the scale of subjective evaluations, it is desirable to refine, under which conditions the evaluated translation can prove to be useful for the reader.

For our purposes the five-or-six-point scale of evaluations will be most appropriate. It makes it possible to evaluate all the essential gradations of the indicator being studied and possesses a sufficient degree of generality (5). To evaluate the clarity of translation we use the following scale of evaluations:

- the sense of the entire text is clear and it causes no doubts;
- the sense of a text is clear, however single sentences require some reflections;
- only overall meaning of a text is clear;
- the overall meaning of a text is understood with difficulty;
- the sense of a text is absolutely unclear.

as a control procedure for the checking of evaluations the following method is used by us. The test subjects underline the unclear to them

places of translation and copy a text, correcting the emphasized places of translation.

The evaluation of the faithfulness of translation to original is also produced by us according to the five-point scale of evaluations.

Evaluations are formulated in the form of these definitions:

- translation transmits very well the sense of original;
- translation transmits well the sense of original;
- translation in its largest part correctly transmits the sense of original;
- translation in its largest part correctly transmits the sense of original;
- translation in no way transmits the sense of original.

When evaluating the acceptability of translations to the test subjects it is proposed to select one of the six versions of the answer:

1. I consider expedient the practical use of such translations in the unedited form on the condition that the system of machine translation to provide for the essential reduction of the periods of the output of signal information.

2. I consider expedient the use of such translations under the condition of a representation along with the translation of a draft.

3. I consider expedient the use of such translations under the condition of a representation along with the translation of the original text and draft.

4. I consider expedient the use of such translations after their insignificant editing.

5. I consider expedient the use of such translations even after their essential editing.

6. I consider inadvisable the use of such translations both in edited and in the unedited form.

In the course of experiments in the evaluation of the quality of translations according to three proposed above parameters were used the standard translations of the texts of the points of the patent formulas from the weekly of "official gazette", executed by the experimental system of Anglo-Russian machine translation of QNIPI by a computer "ural-4". Since in these texts the different types of the errors of translation have been presented in a composite form, it is difficult to judge as to what

degree one or another type of error affects of translation. Therefore in one of the series of experiments in the evaluation of the quality of translation - when evaluating the faithfulness of translation to original - the supplemental procedure has been introduced which makes it possible to judge as to what extent the different types of the errors of translation have been presented in the evaluated text of translation. To the test subjects it is proposed to underline in the text of translation the incorrect translations of terms by two features, grammatical errors by a dotted line and stylistic errors - by a wavy line. This procedure does not give, however, an answer to the question how influences one or another type of error in translation into the general evaluation of quality.

Therefore as text materials specially prepared semi-automatic translations are used by us also. Each of these translations is differed only by one type of evaluation. In experiments the semi-automatic translations of the following nature have been presented:

- 1) the word-by-word translation with solved lexical and grammatical polysemanticism;

- 2) the text with the correct transfer of the grammatical connections

between words, with the exception of the connections between the components of the nominal word combinations being transferred word-by-word;

3) the correctly translated text, but annotating in the case of the lexical polysemanticism of all possible equivalents of the words of the initial language;

4) the correctly translated text, but annotating in the cases of the grammatical homonymy of all possible equivalents of the homonymous forms of the initial language,

5) the correctly translated text, in which part of the words of the initial language has not been transferred, but has been transliterated.

At the beginning of experiments the positive establishment of the test subjects with respect to the machine translation was attained. For this to the test subjects was explained the purpose of works in machine translation, and also the need for ignoring when evaluating the translation of its grammatical and stylistic correctness. The appropriate explanations were introduced into the text of interrogatory forms.

Our experiments were produced by two series. In the first series the texts of semi-automatic and machine translation were presented to the test

subject directly by experimenters. In the second series an experiment was that the test subjects answered the form according to the evaluation of the quality of translation externally. In this form the texts of semi-automatic translation were not included, in it were presented only the texts translated by a computer. Form was published in a periodical the "questions of invention", 1968, No. 5.

At the present time materials on the first series of experiments have been obtained. On the second series of experiments material has been processed partially. ...

In this way it is possible to infer, that to the greatest degree the untranslated vocabulary and to the smallest degree - the destroyed grammatical connections in part of the text make the quality of translations worse. The unsolved grammatical and lexical homonymy makes the translation to a lesser degree worse. Consequently, in order to attain the optimum quality of translation it is necessary under otherwise equal conditions to provide for the system of machine translation with the full dictionary, in which the entire vocabulary of the texts being translated is included. The percentage of the words, not found in a dictionary, should

remain at the lowest level. The translation, intelligible to the reader, can provide for the comparatively simple algorithm which works on the basis of a full dictionary. This is proved by the fact that the text of word-by-word translation and the text of the translation with grammatical agreement, obtained by a computer, have identical evaluations. Exponential in this plan is the fact that the ignoring of the connections between the components of nominal word combinations, and also unsolved grammatical homonymy influence the quality of translation to the smallest degree.

Our conclusion in a somewhat different form has been repeated. With the complication of the algorithms of the syntactic text analysis which have as a goal the detection and the correct transfer of the grammatical connections of the text of original, the degree of the clarity of the translations being obtained substantially does not increase. For practical purposes a simple algorithm is sufficient, but a full dictionary is necessary. SUBJECT CODE 0.

АВТОМАТИЗАЦИЯ ПЕРЕВОДА ТЕКСТОВ

УДК 651.926:681.39

ОЦЕНКА КАЧЕСТВА АВТОМАТИЧЕСКОГО ПЕРЕВОДА

В. А. МОСКОВИЧ, С. М. ВИШНЯКОВА

Целью работы является определение приемлемого и допустимого качества автоматического перевода, рассчитанного как на постредактирование, так и на непосредственное распространение без постредактирования. Такого рода исследование необходимо провести для выяснения возможности практического использования экспериментальных систем автоматического перевода (АП) и для формулировки требований к качеству перевода, который должен обеспечиваться усовершенствованными алгоритмами АП.

Вопрос о допустимом качестве автоматического перевода является в настоящее время одним из центральных вопросов теории и практики АП, решение которого будет во многом способствовать обоснованному прогнозированию дальнейших работ в этой области.

Результаты этих работ по оценке качества перевода не позволяют дать четкого ответа на поставленный в этих работах вопрос: насколько удовлетворительным является автоматический перевод, получаемый в настоящее время? На основании этих работ можно сделать только тот вывод, что тексты, переведенные машинной, по всем характеристикам уступают текстам, переведенным человеком.

В связи с этим выводом отметим наше несогласие с практикой, в упомянутых выше и подобных им работах процедурой предъявления в одних и тех же экспериментах текстов, переведенных людьми, и текстов, переведенных машинной. Дело в том, что при отсутствии предварительной атрибуции перевода к определенному качественному классу трудно судить, о том, что представляет собой текст перевода. До сих пор отсутствует основа для сопоставления качественных характеристик обычного и машинного перевода. Неясно, чем отличается машинный перевод от обычного. Обычный перевод может обладать массой недостатков, которые принципиально могут отсутствовать в машинном переводе (например, гипертрофия черт индивидуального стиля человека-переводчика в переведенном тексте и т. п.). С другой стороны, «гладкость» грамматической структуры обычного перевода (даже при его непонятности) может повлиять на субъективную оценку испытуемых в сторону завышения оценки. Наоборот, шероховатость грамматической структуры, свойственная машинному переводу, но непривычная человеку, может резко снизить оценку перевода при его понятности.

Таким образом, ни одна из предыдущих работ по оценке качества перевода не может быть признана удовлетворительной во всех отношениях. При разработке методики оценки качества переводов патентных текстов необходимо было заново разработать все процедуры экспериментов, учитывая все то рациональное, что предлагалось в работах наших предшественников.

Принципы, положенные в основу разработанной нами методики, сводятся к следующему: 1) четкое разграничение трех независимых параметров измерения качества перевода: понятности перевода, правильности передачи смысла оригинала в переводе и приемлемости перевода в целом; 2) проведение экспериментов с текстами, характеризующихся только каким-нибудь одним типом искажений; 3) отсутствие в экспериментах обычных переводов; 4) нейтрализация факторов, могущих исказить реальную оценку перевода; 5) применение шкал с оптимальным масштабом градации признаков; 6) выработка положительной установки испытуемых к машинному переводу до начала эксперимента.

Выделенные нами параметры измерения качества перевода отличаются относительной независимостью и характеризуют важные аспекты перевода: его понятность, верность оригиналу и его приемлемость. Понятность перевода является основным из оцениваемых параметров качества перевода. Измерить его можно по шкалам субъективных оценок «ясности смысла» перевода.

Второй параметр оценки качества перевода — его верность оригиналу — является в принципе независимым от первого параметра. Он также измеряется на основании субъективных оценок испытуемых.

Наконец, третий параметр — приемлемость перевода — является производным от двух первых. При его измерении, осуществляемом также по шкале субъективных оценок, желательно уточнить, при каких условиях оцениваемый перевод может оказаться полезным читателю.

Для наших целей наиболее подходит пяти-шестибалльная шкала оценок. Она позволяет оценивать все существенные градации изучаемого признака и обладает достаточной степенью общности [5]. Для оценки понятности перевода мы используем следующую шкалу оценок:

- смысл всего текста ясен и не вызывает никаких сомнений;
- смысл текста ясен, однако отдельные предложения требуют некоторых размышлений;
- ясен только общий смысл текста;
- общий смысл текста понимается с трудом;
- смысл текста абсолютно неясен.

В качестве контрольной процедуры для проверки оценок нами применяется следующий прием. Испытуемые подчеркивают неясные им места перевода и переписывают текст, исправляя подчеркнутые места перевода.

Оценка ясности перевода оригиналу также производится нами по пятибалльной шкале оценок. Оценки формулируются в виде таких определений:

- перевод очень хорошо передает смысл оригинала;
- перевод хорошо передает смысл оригинала;
- перевод в большей своей части правильно передает смысл оригинала;
- перевод в большей своей части правильно передает смысл оригинала;
- перевод совершенно не передает смысла оригинала.

При оценке приемлемости переводов испытуемым предлагается выбрать один из шести вариантов ответа:

1. Считаю целесообразным практическое использование таких переводов в неотредактированном виде при условии, что система автоматического перевода обеспечит существенное сокращение сроков выпуска сигнальной информации.

2. Считаю целесообразным использование таких переводов при условии представления наряду с переводом чертежа.

3. Считаю целесообразным использование таких переводов при условии представления наряду с переводом оригинального текста и чертежа.

4. Считаю целесообразным использование таких переводов после их незначительного редактирования.

5. Считаю целесообразным использование таких переводов и после их существенного редактирования.

6. Считаю нецелесообразным использование таких переводов как в отредактированном, так и в неотредактированном виде.

В ходе экспериментов по оценке качества переводов по трем предложенным выше параметрам были использованы стандартные переводы текстов пунктов патентных формул из еженедельника «Official Gazette», выполненные экспериментальной системой англо-русского автоматического перевода ЦНИИПИ на ЭЦВМ «Урал-4». Так как в этих текстах различные типы ошибок перевода представлены в комплексном виде, трудно судить о том, в какой степени тот или иной тип ошибки обуславливает качество перевода. Поэтому в одной из серий экспериментов по оценке качества перевода — при оценке верности перевода оригиналу — введена дополнительная процедура, позволяющая судить о том, в какой мере разные типы ошибок перевода представлены в оцениваемом тексте перевода. Испытуемым предлагается подчеркнуть в тексте перевода неправильные переводы терминов двумя чертами, грамматические ошибки волнистой чертой и стилистические ошибки — пунктирной чертой. Эта процедура не дает, однако, ответа на вопрос, как влияет тот или иной тип ошибки перевода на общую оценку качества.

Поэтому в качестве текстовых материалов нами использованы также специально подготовленные квазиавтоматические переводы. Каждый из этих переводов отличается только одним типом ошибки. В экспериментах представлены квазиавтоматические переводы следующего характера:

1) пословный перевод с решенной лексической и грамматической многозначностью;

2) текст с правильной передачей грамматических связей между словами, за исключением связей между компонентами именных словосочетаний, переводимых пословно;

3) правильно переведенный текст, но с приведением в случае лексической многозначности всех возможных эквивалентов слов исходного языка;

4) правильно переведенный текст, но с приведением в случаях грамматической омонимии всех возможных эквивалентов омонимичных форм исходного языка;

5) правильно переведенный текст, в котором часть слов исходного языка не переведена, а транслитерирована.

К началу экспериментов достигалась положительная установка испытуемых по отношению к машинному переводу. Для этого испытуемым разъяснялось назначение работ по машинному переводу, а также необходимость игнорирования при оценке перевода его грамматической и стилистической правильности. Соответствующие разъяснения были введены в текст опросных анкет.

Наши эксперименты производились двумя сериями. В первой серии тексты квазиавтоматического и автоматического перевода предъявлялись испытуемым непосредственно экспериментаторами. Во второй серии эксперимент состоял в том, что испытуемые отвечали на анкету по оценке качества перевода заочно. В эту анкету не включались тексты квазиавтоматического перевода, в ней были представлены только тексты, переведенные на ЭЦВМ. Анкета была опубликована в журнале «Вопросы изобретательства», 1968, № 6.

К настоящему времени получены материалы по первой серии экспериментов. По второй серии экспериментов материал обработан частично.

Таким образом можно сделать вывод, что в наибольшей степени ухудшают качество переводов непереверенная лексика и в наименьшей степени — разрушенные грамматические связи в части текста. Нерешенная грамматическая и лексическая омонимия ухудшает перевод в меньшей степени. Следовательно, чтобы добиться оптимального качества перевода необходимо при всех прочих равных условиях обеспечить систему автоматического перевода полным словарем, в который включена вся лексика переводимых текстов. Процент слов, не найденных в словаре, должен оставаться на самом низком уровне. Перевод, понятный читателю, может обеспечить сравнительно несложный алгоритм, работающий на основе полного словаря. Это доказывается тем, что текст пословного перевода и текст перевода с грамматическим согласованием, полученный на ЭЦВМ, имеют одинаковые оценки. Показательным в этом плане является и то, что игнорирование связей между компонентами именных словосочетаний, а также неразрешение грамматической омонимии влияют на качество перевода в наименьшей степени.

Повторим наш вывод в несколько иной форме. С усложнением алгоритмов синтаксического анализа текста, имеющих целью обнаружение и правильную передачу грамматических связей текста оригинала, степень понятности получаемых переводов существенно не возрастает. Для практических целей достаточен простой алгоритм, но необходим полный словарь.

SECTION VI

SYSTRAN AND THE CONTROVERSY ABOUT THE STATE OF THE ART

Dispute or controversy about the feasibility of machine translation continues to manifest itself, despite the findings of the 1970-1971 study effort of American linguists and despite the findings of respected Soviet experts. The latest contribution to this controversy appeared in the December, 1971 issue of the journal Science in the form of a letter to the editor from H. W. Sinaiko.

The approach in this letter depends on the writer's offer of statistics in the form of percentages of "errors" in MT -- percentages arrived at by the writer by some unexplained method. The thrust of the letter is certainly negative, even though the writer admits that "it would be unwise to conclude on a less-than-optimistic note because of one set of observations." Nonetheless, Sinaiko's opinion that the "differences between the 1964 and 1971 MT systems were found to be slight and not consistently favoring one or the other system" is certainly a less than optimistic answer to the writer's apparently rhetorical question: "what have we gotten for our money?"

This brusque question is apparently not an inquiry of progress made in the state of the art, but a prelude to the negation of progress in MT. As such, it deserves close examination.

The first point made in the letter is that the same article was translated by a 1964 MT system and by a 1971 system. The article in question was a paper (to be read in Yugoslavia in 1964) written in English, and translated into Russian. No mention is made of mistakes made by the human translator in producing the Russian version. There were, however, errors in the Russian translation--errors which detrimentally affected the computer's retranslation into English. The futility of such an exercise should be considered at this point. MT is designed for practical purposes; it has been developed to provide rapid and accurate translation of primarily scientific texts for scientists who do not read the source language. It has not been meant to serve as an academic merry-go-round of translationalism.

Nonetheless, if humans want to use MT for such academic exercises, MT will serve, if the human-translation input is grammatically and orthographically correct. Unfortunately, this was not the case with the article in question.

Concrete evidence is offered here for the purpose of pointing out the necessity for using genuine Russian material in cases where the machine translation effort is to demonstrate recompensation to the investor. It is also intended to substantiate the claim that SYSTRAN has been developed solely for practical use.

The Sinaiko document was submitted to a native Russian, Dr. P. F. Krasnopolsky, who on the basis of his long teaching

experience in the United States, where he has been dealing with both the Russian and the English languages, is well qualified to judge the use of his native tongue and to determine the instances where it is not used properly or where it does not correspond to the standard literary language in which a native Russian naturally expresses his thoughts on paper. The discrepancies found by Dr. Krasnopol'sky range from incorrect use of Russian (perhaps due to insufficient grammar training), to "Americanisms" and even Russian substandard expressions. For a detailed analysis of the errors found in the Russian text (sixty-five distinct and non-repetitive errors), see Appendix III.

In brief, most of the errors, inadequacies and inaccuracies of the text indicate a poor knowledge of the Russian language. In addition, the Russian text serves as an excellent example of the types of errors human translators may make.

One of the most typical and unavoidable characteristics of human translation is the inevitable tendency towards slight semantic discrepancies. These are inevitable because the human translator unconsciously carries to each text he translates his own prejudices and his own unique system of semantic associations. This, rather than an inadequate knowledge of Russian, may account for the translation from the English "seemed to be . . . slow" (p. 4) into the Russian BYL MEDLEN which means "was slow." Obviously, KAZALS4 MEDLENNYM ("seemed . . . slow") should have been used to achieve a precise

translation. Clearly, that something seems to be the case does not logically imply that it is the case. Yet, in the article in which this imprecise translation occurs, Sinaiko states that with human translators "meaning was preserved with high fidelity." One can only conclude that this optimum condition does not always hold.

Ironically, the abovementioned errors in the human-produced Russian translation of Sinaiko's article actually bear witness to SYSTRAN's sophistication in machine translation. Words which do not exist in the language cannot be expected to be translated, disrupted syntactical structures in the source language cannot yield normal sentences in the target language, and, above all, no computer can infer anything during the structural analysis which is not explicitly expressed in the input. If the input deviates from the standard use of a language, updates must be carried out prior to spinning the tape. Any such deviation forced upon such an operational system as SYSTRAN will jeopardize its capabilities and will only produce adverse effects.

Sinaiko chooses as indices of translation efficiency what he terms "characteristics of MT output," which are "(i) untranslated words and (ii) translated words that have two or more possible meanings in the target language (English in this case)." It is by the use of these indices that the writer nullifies the progress of MT: "The MT (1964) translation

contained 1.2 percent untranslated words and 6.3 percent multiple meanings. The MT (1971) translation contained 2.3 percent untranslated words and 5.3 percent multiple meanings." Unfortunately, he gives no indication of the miscellany aggregate under the term "untranslated words." For example, 38 out of the 1,685 words in the text were not translated in MT 1971; four were not translated because they were misspelled. In other words, the computer correctly rejected human error. Of the remaining 34 untranslated words, the following were counted as untranslated more than once:

VVIDE	(2)
NAXODILS4	(2)
TELEKONFERENQI4	(9)
KOLEBALAS6	(2)

At this point, two facts should be taken into account: (1) as long as people create new vocabulary items (e.g., TELEKONFERENQI4), dictionary updates will be required for their translation. Fortunately, an MT system dictionary can be updated incredibly quickly. (2) Words may remain untranslated by a human translator if he is not equipped to deal with terms in a specialized area. Just as a human translator expert in a specific field, e.g., thermodynamics, is sought to translate a paper in that field, so must an MT system be equipped by dictionary updates (and special provisions for subject areas) to deal with current articles in specified fields.

SYSTRAN is equipped with ten Topical Glossaries to deal with texts in special topical areas. (Telecommunication is not one of these areas.) When it does encounter words it is not equipped to translate, its Not-Found Word Program analyzes the endings of these words so that the output can indicate the part of speech and sentence function of the untranslated word. In other words, the word does not remain simply "untranslated"; it is analyzed as far as this is possible in spite of the absence of a translation equivalent.

Sinaiko's other index of translation efficiency, multiple meanings, also requires a fair amount of discussion. First, it must be pointed out that "multiple meanings", as he uses the term, refers to the appearance in the translated output of a series of words, separated by slashes, any one of which may be chosen as a preferred translation equivalent by the post-editor. In other words, this slashed entry format was used as a convenience for post-editing. This type of translated output is initiated by Subject Code D, which is a special utility of SYSTRAN, rather than an integral feature. When Subject Code D is not selected, only one meaning, the meaning appropriate to the context, will appear in the translated sentence. Other possible meanings for the homograph appear at the end of each sentence in the hexadecimal printout.

Sinaiko also neglected to explain how he arrived at his statistics (5.3 percent multiple meanings). Again, many

occurrences of multiple meanings were counted more than once,
just as the untranslated words were counted more than once:

TRANSLATION/TRANSFER	(27)
EXPERIMENT/EXPERIENCE	(16)
CONNECTION/BOND/RELATION	(9)
RATE/VELOCITY	(8)
POSSIBILITY/CAPABILITY	(4)
REACH/ATTAIN	(4)
AREA/DOMAIN	(2)
COMPRISE/CONSTITUTE	(2)
DIRECTION/TREND	(2)
SEVERAL/SOMEWHAT	(2)

The letter contained two more observations about MT/1971 which further reflected Sinaiko's negative opinion of the progress of MT:

(1) "each of the approximately 80 English sentences had had some editorial modifications, most of them extensive";

(2) "the rate of post-translation editing was slightly less than the rate of human translation."

He does not define "post-translation editing" nor "extensive." Even the most fervid proponents of MT admit the MT produces stylistically rough translations. Any sentient editor, given a free hand, will be anxious to add these stylistic corrections. Editing of this sort might have been extensive, from Sinaiko's viewpoint, especially since the MT translation he saw in March, 1971, was carried out in the midst of a scheduled improvement phase which lasted through the following month. At that time, routines were being rewritten and had not yet been checked out. For this reason,

several L. S. compounds were not recognized, rearrangement did not go into effect, and article insertion was suppressed. In sum, the translation on which he bases his figures was not produced by a fully operating system, but rather by a temporarily disrupted system.

It is difficult to discuss post-editing time requirements because there are various types of post-editing, requiring greater and lesser amounts of time.

For example, a post-editor who checks translated output for clarity of expression in the target language makes very few corrections, and those are only the ones he considers most necessary, if he wishes to forward it to scientists to be read immediately. If the need is less urgent, he may add stylistic changes.

On the other hand, if the post-editor is a bilingual checking the fidelity of the translated output to the source text, then his editing will take at least as much time as he would originally have needed to translate each sentence himself, since he must actually translate the source text to compare it to the target text. However, once a system is in operation, this latter type of post-editing is usually performed only on specific problems in specific sentences, when these problems occur. The sentence by sentence comparison will, of course, be done by the experts who develop the system, as they continually improve MT.

The focal point of this discussion is a paper written in English by Sinaiko and anonymously translated into Russian. Appendices I and II contain the original English text and the translated Russian text. SYSTRAN's translation of the Russian is given below. This translation was done in December, 1971. (Although this translation was done nine (9) months after the text Sinaiko analyzed, it should be logically apparent that the progress made in MT between March and December of 1971 could not exceed the progress made between 1964 and 1971.)

While Sinaiko's translated text makes quite clear the time entailed in human translation, since his letter to Science implies the definite preferability of human translators, LATSEC, INC., employed a professional translator (who has translated professionally from Russian to English since 1963) to translate the same text that was submitted to SYSTRAN. The initial translation was done on a dictaphone and took 100 minutes. After it was typed, the translator then spent 15 minutes "post-editing" his own work. The material was retyped, and then edited to achieve a smoother English text. Three hours had elapsed since the material was first presented to the translator. Needless to say, SYSTRAN's translation was accomplished within minutes. Specifically, SYSTRAN translation of this text took 22.52 seconds (CPU). However, this is not an index of SYSTRAN's standard words-per-minute

translation rate, since a total dictionary look-up is necessary, irrelevant of the size of the text and it was this look-up which consumed most of the time (11.65 seconds) for the translation of this short text.

There were 1,611 words in the text (disregarding punctuation and control words) translated by SYSTRAN. Therefore, SYSTRAN's CPU translation rate for this text was 257,530 Russian words per hour.

SUBJECT CODE 3. Experiments in the field of international teleconferences. *

Kh. Wallace Sinaiko.

Introduction.

Teleconferences refer to the meetings of people who are located far from one another and therefore must use any means for long-distance communication. At teleconferences, who do not have common language can also be the participants. This report sums up some experimental results conducted in the field of the teleconferences which specially relate: a) to the different technical methods of the oral and written translations, capable of being requiring at international teleconferences; and b) to teleconferences, that use the different means of communication. **

* the paper will be presented at the 15th international congress on applied

psychology 2 - 8 August 1964 in Ljubljana, Yugoslavia

** the more comprehensive study, into which these experiments entered, was published G. Bavelas and others in the monograph of the teleconference: the summary of a preliminary study, IDA, study s-138, November 1963. Also about the details of experiments see Kh. U. Sinaiko, the teleconference: the preliminary experiments, the report IDA, p-108, November 1963.

Linguistic experiments.

Background and method. Research on the existing materials gave no indications that in order that in the past practical research on the technical methods used would occur by translators and interpreters of foreign languages. Therefore we conducted a number of experimental observations for the measurement of some characteristic properties of translation and interpretation. We used languages English and French, because they predominate on international conferences and since the highly

skilled linguists who knew both languages, they were available as experimental subjects. The material which translators and interpreters used, it was political in content: the published records of the sessions of the Security Council of united nations. During each experiment we noted the productivity of the linguists from the point of view of accuracy and speed expressed by a number of words per minute. Accuracy was evaluated by the judge-expert who used the criteria of the state department used during the selection of the professional linguists.

Results.

We studied two basic linguistic methods and several of their versions. The first experiments were connected with the generally accepted procedure of full translation. Were used the groups of translators, some from French to English, and others from English to French. One experiment was dedicated to the measurement of the full written translation being produced in the normal way. The results showed, that for full translation much more time was necessary, than for the actual session of united nations. , at the session of united nations for the delivery of speeches one hour of time was necessary namely approximately, whereas our groups of translators spent

37,6 is watch to the full written French or English translation of these same speeches. the full translation rate comprised on the average 5,4 words per minute. The accuracy of full translation was actually irreproachable: on the average it attained 97o/o, and among seven of each of eight excerpts it waverd within the limits between 96o/o and 100o/o.

During another experiment we placed before the groups of translators linguistic problems in reverse order, whereas normally in English-french translators a native language is French, and the group of the translators from French into the English language became accustomed to work only in this direction. When for groups it was necessary to work under uncommon conditions, their accuracy was somewhat lowered, especially in those which basically they spoke in English. (Since the sample of the translation excerpts was small, in this experiment were not calculated the speeds of translations; however, it is possible to say, that the condition the "reverse linguistic problem" had a tendency to slow down the work of translators.).

Was also produced the final, but short experiment which was concerning the technical methods of full written translation. We charged to groups to

translate back into the initial language the samples of the already translated excerpts. Groups worked in the direction of the languages preferred by them. Accuracy was very high: it attained 96-100o/o. Numbers of words per minute were not calculated, but reverse translation into the French language required less time, than the initial translation into English; analogous reverse translation into English was slower than the translation from English into French.

Since for teleconferences the ordinary method of a written translation was inadmissibly slow, we developed a new technical method. it could be used in teleconferences, where the teletype was as the means of communication. We called this method" visual translation". it is produced by a linguist who reads in one language being obtained information and simultaneously dictating the his translation to a typist on a typewriter. Were produced three experiments on visual translation, similar to the those that were used in full written translation. as objects we used oral professional translators, but not written, because this method is based on speed and rapid memorization. The records of the Security Council of united nations were transmitted, both in French and in the English language

with the aid of the teletype equipment at a rate of 60 words per minute.

The first experiment on visual translation showed, that it is much faster than full written translation. For the translation of the one-hour session of united nations it was required 9,7 is watch (21 words per minute). Accuracy was very high: on the average it attained 94o/o, and in two thirds of the total number of excerpts it waverd between 96o/o and 100o/o.

During another experiment we placed before visual translators a reverse linguistic problem. The participating translators said previously, that they will have the bad productivity which will be expressed in a lower speed and the larger number of errors. So it happened: the average accuracy of an American translator was only 84o/o, but of born in France 92o/o. However, the speed of work was lowered only in one translator; another, born in France, translated to the English language just as rapidly as to French.

Finally we conducted one additional short experiment, using visual translation for the reverse translation of material into its initial language. Accuracy was just as high, as in the first experiment on this

type translation. Although the sense was completely preserved, however, not all reversely translated words were expressed identically with their initial form. Among interesting English examples it is possible to mention, that "terrible" became "howling", "shaking" was made "waveing", and "to allow discussion" was turned into "to permit debates". Relative to the speed of reverse visual translation we do not have data.

The test of the means of telecommunication.

Background.

We conducted a number of conferences with the participation of the groups of four and the penta of the man who were located far from one another and had to communicate with the aid of teletype, telephone and in only one instance televisions. With the exception of the television experiment, before conferences was placed the problem to conduct negotiations. The participants conducted the game, during which they had to distribute resources among the group, united by the general purpose, as, for instance, a union, but to every object were given different secret instructions. Every conference was conducted within the framework of some

urgency, i. e. it was necessary to fulfill the group goal in the shortest possible period.

Among fourteen teleconferences which we conducted, there was not one identical from the viewpoint of language, structure of the network or chairman's responsibilities. At some conferences were networks" with a party line" or open: each information being sent by an individual participant, was obtained by all others. Some networks were strongly centralized, where the participant of conferences could turn not to all others, but only to one or to two of them. At some sessions the chairmen were, but not in all cases. We conducted several teleconferences with the group, in which one person spoke only in French, so that it was necessary to use a translator. For the purpose of training we had also several conferences, at which all participants were present face to face.

Results.

The participants of conferences preferred to negotiate with the aid of the most impersonal means of communication. Teletype was the favored method of the conducting of negotiations, and by least popular conferences were face to face. A television and a telephone were acknowledged by the

intermediates between these extremes. The participants of conferences said, that they have a possibility of manifesting much greater intractability, using teletype, than with all other methods.

For the conducting of negotiations at teleconferences teletype allowed the determined advantages single in its kind. First, it was very useful to have a printed record of all informations. In the second place, the need for composing informations for a transfer on teletype assigned on the participants of conferences a known discipline, which does not always relate to direct conversational communication. (However, despite the more compressed language, accepted at teletype conferences, at these conferences of agreement they were attained much more slow, than at sessions face to face or on a telephone). Thirdly, on some teletype networks there was the present possibility of a bilateral and simultaneous transfer (for example A could send an information to v at the same time that to the v it sent its notice to A). This possibility had advantages as compared with the other methods, but it also created confusion, disrupting the normal sequence of negotiations. Fourthly, in the case of the centralized teletype network the picture which indicated the great potential power of its chairman was

obtained. It had an absolute control over who with whom was found in telecommunication, and at its station could edit the passed informations, them detain or change order. Fifthly, since teletype does not have a mechanism for feedback for the checking the fact that an addressee actually obtained or understood sent to it an information, that for the achievement of this it is necessary to use a supplemental procedure. In the negotiations with the aid of the oral communication of confirmation they are transmitted in the form short answers; at sessions the non-verbal confirmations in the form of the mimicry or gestures are face to face also. Sixthly, a skillful translator was able bilaterally to visually translate teletype informations (from English to French and vice versa), which corrected the disadvantageous position of the non-English-speaking participant of conference as compared with others, that led negotiations in the English language. Finally, it is necessary to note, that at the teleconferences equipped both teletype and by telephone communication, the participants rarely used the latter.

Our short experiment on the utilization of a television at conferences detected no special advantages in the fact that the persons the leading

negotiations and far away, can see being located each other. On the contrary, in the opinion of the participants of these sessions their own mimicry and the other non-verbal manifestations only contribute the delay of the work of conference, having distracted the attention of viewers. The participants of conference also did not approve of the television equipment, in which they could see not only those, with whom they led an interview, but also themselves.

Conclusion.

In the coming years will be reached the large technical successes which will permit the groups of the far being located people to confer with each other. In order to make such teleconferences effective, it is necessary to learn much which concerns how to use these technical capabilities with the greatest benefit for humanity. We represented a resume about the last preliminary experiments relating to two little studied factors in connection with the conferences at a distance, namely: to the linguistic profession and means of communication. The just now obtained experimental results which were here mentioned, they can be as

02-12-72

PAGE 12

guiding landmarks for the those for whom in the near future it is necessary to develop the procedures of teleconferences. Furthermore, we think, that we succeeded in visually confirming the practicability of the experiments dedicated to the variable amounts of teleconferences. we hope, that the other interested ucenye-bixevioristy will study by these problems tied with them. It is completely possible, that on one fine day such a meeting, as today international congress, will be conducted in the form teleconference and each of us will sit at its writing desk in its own country.

SECTION VII

SYSTRAN'S POSITION

A grammar can be used to teach languages or to analyze them. The use of a machine grammar for translation has one main goal: to free the human intellect from tedious, mechanical chores in order that it may pursue its proper vocation--creativity.

The purpose of scientific and technical translation is to facilitate the rapid transmittal of new developments in human knowledge from one country or one culture to another. Uniform translation overcomes language barriers and individual idiosyncratic preconceptions. Because of this, it can be a valuable contributor to mutual understanding and an impetus to mutual aid in development through the sharing of ideas. These two conditions are probably among the most important prerequisites to world peace.

Uniform translation can best be achieved by machine rather than by humans for several reasons. First, a computer has no "native" language. It can translate equally well from language A to language B, from B to A, or from X to Z. Second, a computer has no preconceptions or cultural biases regarding any language it deals with, or regarding the people who speak the language. Third, it hasn't the educational limits of humans. It can store almost unlimited information from any

number of subject fields.

These three factors must inevitably play a large part in producing unambiguous information, free from bias, prejudice or preconception. A grammar which is instrumental in producing unambiguous information by overcoming language barriers can be regarded as a definitively positive asset in language science. Such a grammar is the SYSTRAN grammar.

SYSTRAN, like any grammatical theory, must be tested for adequacy. If it does not adequately account for the phenomena it purports to, it should be modified or discarded. Many linguists have apparently decided that the most efficient way to test a grammar is to use computers (cf. Woods, Kuno, Hays, etc.).

The unbiased computer provides the best evaluation of a grammar because it blindly carries out any and all operations specified or allowed by the grammar, delivering up, in its output, all possible consequences of the theory programmed into it. The computer is the presiding master of logic.

SYSTRAN was designed to be used by the computer for translation purposes. Therefore, it is a recognition grammar. As it is used again and again for computer translations, it is constantly tested for its adequacy in accurately transmitting information from Russian to English with fidelity and intelligibility. As a result of this constant testing, SYSTRAN is continuously modified and improved.

SYSTRAN, as an automatic recognition grammar designed specifically for computer use, must be supplied with a

sufficient number of well-organized grammatical and semantic codes in order that it may carry out its function. These codes must be designed in response to the capabilities and requirements of the computer.

The most obvious example of grammatical coding specifically designed for the computer is SYSTRAN's approach to part of speech categorization. SYSTRAN has two major divisions: primary parts of speech and secondary parts of speech. Primary and secondary part of speech assignment is oriented towards the fullest utilization of the categorization. For example, the Russian pronouns EGO, EE, and IX function either as personal pronouns or as possessive pronouns. In many instances the three function in the same way in the source language. Therefore, their primary part of speech (POS) is POS78.

When the target language equivalent is chosen, it becomes necessary to specify which of the three is to be translated. For this purpose, secondary (BPQ) parts of speech are assigned (7B=EE, 7C=EGO, 7D=IX).

The design of the SYSTRAN grammar also allows codes to express special types of relationships extending over several clauses. For example, the ESLI-TO or KOGDA-TO logical relationship can be coded. If this were implemented, when SYSTRAN encountered either a ESLI or KOGDA clause, it would simultaneously be alerted to expect a TO clause later in the sentence. It then could use this type of information in establishing clause levels within the sentence discussed earlier in this

report (II, "A Comparison of Human and SYSTRAN Parsing"). These codes, if implemented, could have shown that the ESLI clause is logically related to the preceding clause in the sentence quoted from Iordanskaja (see Figures 3 and 4). In implementing codes of this sort, it first must be determined what linguistic analysis is most appropriate. For example, in this case, the linguistic analysis must determine if ESLI is logically related to the preceding predicate in the KOTOROE clause, or if its relationship is with the main clause.

The point to be made here is that SYSTRAN is ready to devise and implement codes of this sort, as soon as SYSTRAN's data gathering feature compiles a sufficient number of sentences to allow all these and similar problems to be solved. These data can then be analyzed by persons trained in SYSTRAN grammar concepts. The result of the analysis can be incorporated into SYSTRAN, thereby further improving the System.

SECTION VIII

THE FUTURE OF MACHINE TRANSLATION

As recently as 1964, Eugene Nida remarked that "despite the enormous storage capacity in modern computers, they are nevertheless incredibly less complicated than the human brain, which has approximately 10^{14} cells."¹ In equating complexity with the number of cells in the human brain, Nida implied that the computer was less complex because it had fewer "cells." Less than a decade later, a 1971 prognostication of the development of the computer predicted "radical changes . . . in the area of mass storage technology. Whichever methods--lasers, holography, magnetic bubbles, etc.--are used, they promise greater access speeds and vast storage capacities--as great as 10^{13} or 10^{14} bits."²

In effect, computers will soon have memory capacities equivalent to those of the human brain. Such an eventuality requires system designs able to make the most efficient use of the resources soon to be at our disposal. SYSTRAN's open-ended design anticipates these developments.

¹Eugene A. Nida, Toward a Science of Translating, Leiden, 1964, p. 263.

²"A Quarter Century Look Ahead at the Computer," Communications of the ACM, Vol. 14, No. 9, Sept., 1971, p. 615.

In the following, a few examples will be given to show how the SYSTRAN concept can use such expanded memory capacities to improve the quality of machine translation. This concept has already been proven to be valid not only for Russian-English translation but for translation between other source and target languages.

The first step will be the development of a much more detailed topical glossarization of the individual entry level and of the compound level. The compound concept will be expanded to include interrupted compounds, thus taking into consideration the mutual interlocking influence of several words irrelevant of the position of occurrence.

Presently, an operationally oriented semantic categorization is being developed within SYSTRAN. Such categorization makes it possible to introduce general types of multiple-meaning resolutions by determining how semantic categories carry out their influence within definable syntactic connections. In the following sentence segments, the verb RASTI occurs with two different meanings: INCREASE and GROW. The choice of either meaning is dependent upon the subject's carrying semantic code: "animate." In the first example, the subject TEMPERATURA does not have the "animate" code; consequently the translation is INCREASE. In the second example, the subject BAKTERII carries the "animate" code; therefore the translation is GROW.

- (a) . . . A TEMPERATURA GAZA BYSTRO RASTET PRI
UVELICENII SKOROSTI.

. . .BUT THE TEMPERATURE OF THE GAS RAPIDLY
INCREASES AT THE INCREASE OF VELOCITY.

(b) . . . I OKRUJAHSEI SREDY, V KOTOROI RASTUT
BAKTERII . . .

. . . AND OF SURROUNDING MEDIUM, IN WHICH
BACTERIA GROW . . .

The SYSTRAN dictionary system can accomodate, in variable length format after every dictionary entry, up to 2,048 semantic codes, expressing similar types of semantic categorization. In addition to recognizing and using semantic categories for multiple meaning resolution, SYSTRAN also makes use of contextual information and has inherent capabilities for resolving semantic ambiguities on the contextual level. A typical example of SYSTRAN's ability to retain contextual information is the retention of human and animate codes after each sentence is analyzed.

Such codes are retained for the last sentence's subject or for its last object--or for a pronoun if it occurred at the end of the sentence. The code is placed into analysis area. This area is available before the main analysis of the sentence. If it is needed, the information in the 160 bytes reserved for a hypothetical zero word is not cleared when a new sentence analysis begins. The indication not to clear the information in these bytes is given by the "remembering" routine whenever there is a necessity to transfer information from one sentence into another. With the help of this contextual feature, SYSTRAN can easily decide that, in a specific sentence with two identical

pronouns, one EE refers to an inanimate noun, and that its translation should be ITS, while the second pronoun should be HER.

Once a pronoun has occurred, and semantic codes from its noun antecedent have been transferred to it, this pronoun can serve as an originator for these memory codes.

SYSTRAN is also open for the introduction of a systematic remembering feature for nouns in order to improve the selection of definite articles during translation. Each time a new noun occurs with an article, it will be remembered. In addition, a remembering count can be introduced, which will allow the establishment of "weighting" criteria. In other words, previous occurrences of nouns under analysis can be "remembered" to determine how often it occurred with definite articles and how often it occurred with indefinite articles, or without articles.

To further contextual analysis, SYSTRAN can easily be modified to retain whole paragraphs or even larger text segments in the high speed memory in computer analyzed form. With only 150,000 bytes of additional core, up to 50 sentences can be kept in the memory where there is immediate access to every dictionary code in the source or target language. Moreover, each word in these fifty sentences would carry its codes in the computer intermediate language.

The following examples demonstrate the sorts of distinctions SYSTRAN is potentially able to make in choosing an appropriate

translation out of several possible meanings for a given lexical
item:

ESTESTVENNO ZAKLHCIT6, CTO SLO1 NA VENERE SOSTOIT
IZ LED4NYX CASTIQ

ZAKLHCIT6 = CONCLUDE

DIAPAZON PERIODOV MORSKIX VOLN ZAKLHCEN MEJDU 8 I
12 SEK

ZAKLHCIT6 = INCLUDE

VYDELENNYE NAMI PARAMETRY IZMERENI4 KACESTVA PEREVODA
OTLICAHTS4 OTNOSITEL6NO1 NEZAVISIMOST6H

VYDELENNY1 = SELECTED

RASSMATRIVAETS4 PROBLEMA PROISXOJDENI4 PRODOL6NYX VOLN,
VYDELENNYX RANEE V MIKROSELSMAX TIXIX RALONOV SSSR

VYDELENNYX = PICKED UP

TEPLOTA VYDELENN4 V 4DERNOM VZRYVE NAMNOGO BOL6WE CEM
3KVIVALENTNA4 TEPLOTA VSEX GIDRO3LEKTRICESKIX STANQ11

VYDELENN4 = RELEASED

BIBLIOGRAPHY

- "A Quarter Century Look Ahead at the Computer," Communications of the ACM, Vol. 14, No. 9, Sept., 1971.
- Dykema, Karl W. "Where Our Grammar Came From," Readings in Applied English Linguistics, H. B. Allen, ed., New York, 1958.
- Final Report, "Feasibility Study on Fully Automatic High Quality Translation," University of Texas at Austin, Contract F30602-70-C-0129.
- Iordanskaja, L. N. "Avtomaticeskij sintaksiceskij analiz," Kibernetika v monografijax, 2, izd. Nauka, Novosibirsk, 1967, pp. 49-74.
- Nida, Eugene A. Toward a Science of Translating, Leiden, 1964.
- Sevbo, I. P. "O gromozdkosti sintaksiceskix structur," Naucno-Texniceskaja Informacija: Automatizacija perevoda tekstov, ser. 2, No. 2., Moscow, 1971, pp42ff.
- Sinaiko, H. W. "Letter to the Editor," Science, Vol. 174, No. 4015, 17 December, 1971.

APPENDIX I

EXPERIMENTS IN INTERNATIONAL TELECONFERENCING*

H. Wallace Sinaiko,

Introduction

Teleconferencing refers to meetings among people who are remote from one another and, therefore, must use some means of telecommunication. There may also be some participants in teleconferences who do not all share a common language. This report summarizes some of the results of experiments on teleconferencing, particularly those having to do with: a) various language interpretation and translation techniques that might be required in international teleconferencing; and, b) teleconferences using different media of communication.**

Language Experiments

Background and Method. Background studies failed to show any previous experimental research on the techniques used by language translators or interpreters. Therefore, we ran a series of pilot studies to measure some attributes of translation and interpretation. We used English and French because

* Paper to be presented at SV International Congress of Applied Psychology, August 2 - 8, 1964; Ljubljana, Yugoslavia.

** The larger study of which these experiments were a part has been published in A. Bavelas, et al Teleconferencing: Summary of a Preliminary Study, IDA Study S-138, November 1963. Also for more detail on the experiments see H. W. Sinaiko, Teleconferencing: Preliminary Experiments, IDA Research Paper, P-108; November 1963.

of the prominence of these languages in international conferences and because highly skilled linguists who knew both languages were readily available as experimental subjects. The material handled by the translators and interpreters was political in content: published minutes of United Nations Security Council meetings. In each of the experiments we scored performance of the linguists for speed -- expressed below as a words-per-minute rate -- and accuracy. Accuracy was rated by a single expert judge who followed the Department of State criteria used to select professional linguists.

Results. We studied two basic language techniques and several variations of each. The first experiments were concerned with conventional, full translation techniques. Teams of translators were used -- one doing French-to-English and the other English-to-French. One experiment was intended to measure full translation as it is done traditionally. Results showed that the time required to do full translation was very much longer than the original UN meeting. That is, in the UN meeting the speeches were delivered in about one hour; our translation teams spent 37.6 hours to fully translate the same speeches into written French or English. The full translation rate averaged 5.4 words per minute. The accuracy of full translation was virtually perfect: average ratings were 97% and seven-eighths of all passages were scored between 96% and 100%.

In another language experiment we reversed the target languages of the translation teams. That is, the English-to-French

team had translators for whom French was the native tongue and the French-to-English team was accustomed to working only in that direction. When the teams were required to work in these unaccustomed directions there was some loss of accuracy, particularly for the English-speaking team. (Because the sample of passages translated was small, no translation rates were computed for this experiment; we can report, however, that the "reversed target language" condition tended to slow down the translators.)

A final brief experiment was done with the full translation technique. We had the teams re-translate back into the original languages samples of already-translated passages. Teams worked in the direction of their preferred target languages. Accuracy was very high: all passages were scored 96% to 100%. No word rates were computed but re-translation back into French took less time than the original translation into English; similarly, retranslation back into English was slower than the English to French translations.

Because conventional full translation seemed to be unacceptably slow for teleconferences in which the medium of communication between participants was teletype. We have called the technique "sight translation." It is done by a linguist who reads an incoming message in one language and dictates a simultaneous translation to a typist. Three experiments, paralleling those for the full translation technique, were done on sight translation. We used professional interpreters as

subjects rather than translators because the technique emphasizes speed and short term memory. The UN Security Council minutes were transmitted via 60-word-per-minute teletype equipment, both in French and English.

The first experiment on sight translation showed it to be much faster than full translation. The one hour UN meeting required 9.7 hours to be translated (21 words per minute). Accuracy was very high: the average score was 94% and two-thirds of all passages were rated 96% to 100%.

We reversed the target languages of the sight translators in another experiment. The interpreter-subjects themselves predicted that they would perform poorly in terms of slower rates and more errors. This was, in fact, what happened: average accuracy scores were 84% and 92% for the American and native French interpreter, respectively. However, translation rates went down only in the case of one interpreter; the other--the native French subject--worked as fast into English as he had into French.

Finally, we conducted one short experiment using sight translation to re-translate material back into its original language. Accuracy scores were as high as in the first sight translation experiment. Although meaning was preserved with high fidelity, not all words were re-translated into their original identical form. Some interesting English examples of these changes are: "flagrant" became "blatant," "tottering" became "vacillating," and "to allow discussion" became "to

permit debate." We have no data on the speed of sight re-translation.

Telecommunications Media Experiments

Background. We ran a series of conferences among groups of four and five people who were remote from one another and who had to communicate using teletype, telephone, or, in one case, television. With the exception of television, these conferences involved a negotiation task. Participants played a game in which they had to allocate resources to a common group goal, as in an alliance, but with individual differences in the secret instructions given to each subject. Each conference was conducted with some urgency; that is, the group goal was to be met as soon as possible.

In the fourteen teleconferences we ran no two were identical in terms of network structure, chairmanship role, or language. Some conferences had a "party-line" or open network: every message sent by a participant was received by all the others. Some networks were highly centralized and permitted a conferee to address himself to one or two, but not all, of the others. There were conference chairmen in some meetings but not all. We ran several teleconferences with one French-speaking subject in the group so that an interpreter had to be used. We also ran a few face-to-face practice conferences.

Results. Conferees preferred to bargain using the most impersonal communications medium. Teletype was most preferred as a negotiation medium and face-to-face meetings were least popular. Television and telephone were assessed as between these extremes. Conferees said they could take much tougher positions using the teletype than with any of the other techniques.

Teletype offered certain unique advantages to negotiation teleconferences. First, the accumulated printed record of all messages was useful. Second, the need to compose messages for teletype transmission imposed a discipline on the conferees not necessary in direct voice communication. (However, in spite of the generally more terse language used in the teletype conferences, these meetings were much slower in reaching solutions than the face-to-face or telephone meetings.) Third, some teletype networks provided true, two-way simultaneous transmission (e.g., A could send a message to B at the same time B was sending a message to A). This capability was an advantage over other media but it also created confusion by disturbing the normal sequence of bargaining. Fourth, the presence of a chairman in a centralized teletype network demonstrated the tremendous potential power of such an individual. Teletype chairmen had absolute control over who communicated with whom; the chairmen could also edit, delete, or change the sequence of messages going through their stations. Fifth, because teletype has no built-in feedback mechanism to insure

that the addressee of a message has actually received or understood what was sent to him, this must be done procedurally. Acknowledgments in voice-mediated discussions are easily transmitted by brief answers; face-to-face meetings also include non-verbal acknowledgments such as facial expressions or gestures. Sixth, a skillful interpreter was able to handle two-way sight translations of teletype messages (English-to-French and vice versa) which put a non-English-speaking conferee at no obvious disadvantage from other conferees who negotiated in English. Finally, in teleconferences with both teletype and telephone available to them, conferees rarely used the telephone facilities.

Our brief experience with television as a conference medium showed no particular advantage in having remote negotiators see each other. On the contrary, the participants in this meeting felt that their own facial expressions or other non-verbal behavior might have hurt the progress of the session by distracting the viewer. Also, the conferees did not like a television arrangement in which they could see themselves as well as the other parties with whom they were talking.

Conclusion

The next few years will bring great technical advances enabling groups of people who are physically remote to confer with one another. If such teleconferences are to be effective

a good deal needs to be learned about how to arrange such technical capabilities for optimum human use. We have summarized some results of recent preliminary experiments on two areas about which little is known in the context of remote conferencing: language services and communications media. The immediate tentative results reported here might provide some guides to people who have to arrange teleconference procedures in the near future. Beyond this, we believe we have demonstrated the feasibility of experimentation on teleconferencing variables. We hope that other interested behavioral scientists will address themselves toward these and related problems. It is not inconceivable that some day a meeting, such as today's International Congress, will be conducted as a teleconference with each of us in his own country at his own desk.

APPENDIX II

ОПЫТЫ В ОБЛАСТИ МЕЖДУНАРОДНЫХ
ТЕЛЕСОВЕЩАНИЙ^О

Х. Уоллес Синайко

Введение

Телесовещания имеют отношение к встречам людей, которые находятся вдали друг от друга и поэтому должны пользоваться какими-то средствами для дальней связи. На телеконференциях могут также оказаться участники, не имеющие общего языка. Этот доклад резюмирует некоторые результаты опытов, проведенных в области телесовещаний, которые специально относятся:

а/ к различным техническим методам устных и письменных переводов, могущих понадобиться на международных телесовещаниях; и в/ к телеконференциям, пользующимся различными средствами связи.^{ОО}

^О Доклад будет представлен на 15-ом Международном Конгрессе по Прикладной Психологии 2 - 8 Августа 1964 г. в Любляне, Югославии.

^{ОО} Более подробное исследование, в которое вошли эти опыты, было опубликовано г. Бавелас и др. в монографии Телесовещания: Резюме предварительного исследования, IDA, Исследование S - 138, Ноябрь 1963 г. Также о подробностях опытов см. Х. У. Синайко, Телесовещания: Предварительные опыты, Доклад IDA, P - 108, Ноябрь 1963 г.

Лингвистические опыты

Предпосылки и Метод. Изучение существующих материалов не дало никаких указаний на то, чтобы в прошлом происходило практическое изучение технических методов, применяемых письменными и устными переводчиками иностранных языков. Поэтому мы провели ряд опытных наблюдений для изменения некоторых характерных особенностей письменного и устного перевода. Мы пользовались языками английским и французским, потому что они преобладают на международных конференциях и т.к. высококвалифицированные лингвисты, знавшие оба языка, были в нашем распоряжении в качестве экспериментальных объектов. Материал, которым пользовались письменные и устные переводчики, был политического содержания: опубликованные протоколы заседаний Совета Безопасности Объединенных Наций. Во время каждого опыта мы отмечали производительности лингвистов с точки зрения точности и скорости, выраженной числом слов в минуту. Точность оценивалась судьей-экспертом, который пользовался критериями Государственного Департамента, применяемыми при отборе профессиональных лингвистов.

Результаты. Мы изучили два основных лингвистических метода и несколько их вариантов. Первые опыты были связаны с общепринятой методикой полного перевода. Были использованы группы переводчиков, один с французского на английский, а другие с английского на французский. Один опыт был посвящен измерению полного письменного перевода, производимого обычным путем. Результаты показали, что для полного перевода требовалось

гораздо больше времени, чем для самого заседания Объединенных Наций. А именно, на заседании Объединенных Наций для произнесения речей требовался приблизительно один час времени, тогда как наши группы переводчиков расходовали 37,6 часов на полный письменный французский или английский перевод этих самых речей. Полная скорость перевода составляла в среднем 5,4 слов в минуту. Точность полного перевода была фактически безупречной: в среднем она достигала 97 о/о, а среди семи из каждых восьми отрывков она колебалась в пределах между 96 о/о и 100 о/о.

Во время другого опыта мы поставили перед группами переводчиков лингвистические задачи в обратном порядке, тогда как нормально у англо-французских переводчиков родным языком является французский, а группа переводчиков с французского на английский, а группа переводчиков с французского на английский язык привыкла работать только в этом направлении. Когда группам пришлось работать в необычных условиях, их точность несколько понизилась, особенно у тех, которые в основном говорили по-английски. /Т.к. образец переведенных отрывков был мал, в этом опыте не были подсчитаны скорости переводов; однако, можно сказать, что условие "обратной лингвистической задачи" имело тенденцию замедлить работу переводчиков./

Был также произведен окончательный, но короткий опыт, касавшийся технических методов полного письменного перевода. Мы поручили группам перевести обратно на первоначальный язык образцы уже переведенных отрывков. Группы работали в направлении ими предпочитаемых языков. Точность была очень

высокой: она достигала 96 - 100 о/о. Числа слов в минуту не были подсчитаны, но обратный перевод на французский язык потребовал меньше времени, чем первоначальный перевод на английский; аналогичный обратный перевод на английский был медленнее, чем перевод с английского на французский.

Т.к. для телесовещаний обычный способ письменного перевода был недопустимо медлен, мы разработали новый технический метод. Он мог применяться на телеконференциях, где телетайп служил средством связи. Мы назвали этот метод "зрительным переводом". Он производится лингвистом, читающим на одном языке получаемое сообщение и одновременно диктующим его перевод переписчику на пишущей машинке. Было произведено три опыта со зрительным переводом, подобных тем, которые применялись при полном письменном переводе. В качестве объектов мы пользовались устными переводчиками-профессионалами, а не письменными, потому что этот метод основан на скорости и быстром запоминании. Протоколы Совета Безопасности Объединенных Наций были переданы, как на французском, так и на английском языке с помощью телетайпного оборудования со скоростью 60 слов в минуту.

Первый опыт со зрительным переводом показал, что он гораздо скорее, чем полный письменный перевод. Для перевода одночасового заседания Объединенных Наций потребовалось 9,7 часов /21 слово в минуту/. Точность была очень высокой; в среднем она достигала 94 о/о, а у двух третей общего числа отрывков она колебалась между 96 о/о и 100 о/о.

Во время другого опыта мы поставили перед зрительными переводчиками обратную лингвистическую задачу. Участвовавшие переводчики скажали заранее, что у них будет плохая производительность, которая выразится в меньшей скорости и большем числе ошибок. Так оно и случилось: средняя точность американского переводчика была лишь 84 о/о, а у родившегося во Франции 92 о/о. Однако, скорость работы понизилась только у одного переводчика; другой, родившийся во Франции, переводил на английский язык столь же быстро, как и на французский.

Напоследок мы провели еще один короткий эксперимент, пользуясь зрительным переводом для обратного перевода материала на его первоначальный язык. Точность была столь же высокой, как и в первом опыте с переводом этого типа. Хотя смысл был полностью сохранен, однако, не все обратно переведенные слова были выражены идентично с их первоначальной формой. Среди интересных английских примеров можно упомянуть, что "ужасный" стал "вопиющим", "трясущийся" сделался "колеблющимся", а "позволить дискуссии" обратилось в "разрешить дебаты". Относительно скорость обратного зрительного перевода у нас нет данных.

Испытание средств телесвязи

Предпосылки. Мы провели ряд конференций при участии групп из четырех и пяти человек, которые находились вдали друг от друга и должны были сообщаться между собой с помощью телетайпа, телефона и в одном только случае телевидения.

За исключением опыта с телевидением, перед совещаниями была поставлена задача вести переговоры. Участники проводили игру, во время которой они должны были распределить ресурсы среди группы, объединенной общей целью, как например союзом, но каждому объекту были даны различные секретные инструкции. Каждая конференция проводилась в рамках некоторой безотлагательности, т.е. требовалось удовлетворить групповую цель в возможно кратчайший срок.

Среди четырнадцати телеконференций, которые мы провели, не было ни одной идентичной с точки зрения языка, структуры сети или обязанностей председателя. На некоторых конференциях были сети "с партийной линией" или же открытые: каждое сообщение, посылаемое отдельным участником, получалось всеми остальными. Некоторые сети были сильно централизованы, где участник совещаний мог обращаться не ко всем остальным, а только к одному или двум из них. На некоторых заседаниях бывали председатели, но не во всех случаях. Мы провели несколько телеконференций с группой, в которой один человек говорил только по-французски, так что пришлось пользоваться переводчиком. С целью тренировки мы имели также несколько конференций, на которых все участники присутствовали лицом к лицу.

Результаты. Участники конференций предпочитали договариваться с помощью наиболее безличных средств связи. Телетайп был излюбленным способом ведения переговоров, а наименее популярными были совещания лицом к лицу. Телевидение и телефон были признаны промежуточными между этими крайностями. Участники конференций

говорили, что они имеют возможность проявлять гораздо большую несговорчивость, пользуясь телетайпом, чем при всех иных методах.

Для ведения переговоров на телеконференциях телетайп предоставлял определенные и единственные в своем роде преимущества. Во-первых, было весьма полезно иметь печатную запись всех сообщений. Во-вторых, необходимость составлять сообщения для передачи по телетайпу налагала на участников конференций известную дисциплину, что не всегда относится к прямой разговорной связи, /Однако, несмотря на более сжатый язык, обычно применяемый на телетайпных конференциях, на этих совещаниях соглашения достигались гораздо медленнее, чем на заседаниях лицом к лицу или по телефону/. В-третьих, на некоторых телетайпных сетях существовала настоящая возможность двухсторонней и одновременной передачи /например, А мог посылать сообщение по адресу В в то самое время, когда В высылал свое извещение по адресу А/. Эта возможность имела преимущества по сравнению с другими способами, но она также создала замешательство, нарушая нормальную последовательность переговоров. В-четвертых, в случае централизованной телетайпной сети получалась картина, указывавшая на громадную потенциальную власть ее председателя. Он имел абсолютный контроль над тем, кто с кем находился в телесвязи, и на своей станции мог редактировать проходившие сообщения, их задерживать или же менять очередность. В-пятых, так как телетайп не имеет механизма обратной связи для проверки того, что адресат действительно получил или понял посланное ему сообщение, то

для достижения этого приходится применять дополнительную процедуру. При переговорах с помощью голосовой связи подтверждения передаются в виде коротких ответов; при заседаниях лицом к лицу имеются также бессловесные подтверждения в виде мимики или жестов. В-шестых, искусный переводчик был в состоянии двусторонне зрительно переводить телетайпные сообщения /с английского на французский и наоборот/, что исправляло невыгодное положение не говорившего по-английски участника конференции по сравнению с другими, ведущими переговоры на английском языке. Наконец, надо отметить, что на телеконференциях, оборудованных как телетайпной, так и телефонной связью, участники редко пользовались последней.

Наш короткий опыт с применением телевизии на конференциях не обнаружил никаких особых преимуществ в том, что лица ведущие переговоры и находящиеся вдалеке, могут видеть друг друга. Наоборот, по мнению участников этих заседаний их собственная мимика и другие бессловесные проявления только способствовали задержке работы совещания, отвлекая внимание смотрящих. Участники конференции также не одобрили телевизионное оборудование, в котором они могли видеть не только тех лиц, с которыми они вели собеседование, но и самих себя.

Заключение

В течение ближайшего ряда лет будут достигнуты большие технические успехи, которые предоставят возможность группам вдали

находящихся людей совещаться друг с другом. Чтобы сделать такие телеконференции эффективными, придется узнать многое, касающееся того, как использовать эти технические возможности с наибольшей пользой для человечества. Мы представили резюме о последних предварительных опытах, относящихся к двум мало изученным факторам в связи с совещаниями на расстоянии, а именно: к лингвистической службе и средствам связи. Только-то полученные экспериментальные результаты, которые здесь были упомянуты, могут служить путеводными вехами для лиц, которым в ближайшем будущем придется разрабатывать процедуры телеконференций. Кроме того, мы думаем, что нам удалось наглядно подтвердить осуществимость опытов, посвященных переменным величинам телесовещаний. Мы надеемся, что другие заинтересованные ученые-бихевиористы займутся этими и связанными с ними проблемами. Вполне возможно, что в один прекрасный день такое собрание, как сегодняшний Международный Конгресс, будет проводиться в виде телеконференции и каждый из нас будет сидеть за своим письменным столом в своей собственной стране.

APPENDIX III

ANALYSIS OF "EXPERIMENTS IN INTERNATIONAL TELECONFERENCING"

H. W. Sinaiko

Analyzed by Dr. Paul Krasnopolsky

Page 1

ENGLISH ORIGINAL	CORRECT TRANSLATION	WAS TRANSLATED	NOTE
Teleconferencing refers to meetings among people	Телесовещания относятся к <u>совещаниям</u> людей...	Телесовещания имеют отношение к <u>встречам</u> людей...	"Встреча" means "meeting face-to-face"
...experiments... particularly those having to do with...	...опытов... в <u>особенности тех, которые имеют дело с...</u>	...опытов... которые <u>специально относятся к...</u>	inaccurate translation
...translation techniques that might be required...	с техническими методами, <u>которые могут понадобиться</u>	...техническим методам... переводов, <u>могущих понадобиться...</u>	"могущий" - this participle is not used now
The larger study...	Более обширное исследование...	Более <u>подробное</u> исследование...	inexact translation

ENGLISH ORIGINAL	CORRECT TRANSLATION	WAS TRANSLATED	NOTE
...has been published in A. Bavelas, et al. <u>Teleconferencing</u> :	..было опубликовано в монографии Г. Бавеласа и др.: "Телесовещания"...	..было опубликовано Г. Бавелас и др. в монографии <u>Телесовещания</u> :....	Word order!
Background studies...	Изучение литературы <u>вопроса</u> ...	Изучение <u>существующих материалов</u> ...	inaccurate translation
...failed to show any previous experimental research on the techniques...	не дало никаких указаний на наличие в прошлом какого-либо экспериментального исследования технических методов...	не дало никаких указаний на то, чтобы в прошлом происходило практическое изучение технических методов...	inexact translation
...a series of pilot studies...	ряд <u>экспериментальных исследований</u> ...	ряд <u>опытных наблюдений</u> ...	"наблюдение" means observation
...to measure some attributes of... translation	для <u>определения</u> <u>некоторых</u> <u>особенностей</u> <u>перевода</u> ...	для <u>измерения</u> <u>некоторых</u> <u>характерных особенностей</u> <u>перевода</u> ...	inappropriate word

ENGLISH ORIGINAL	CORRECT TRANSLATION	WAS TRANSLATED	NOTE
..we scored performance of the linguists	Мы отмечали <u>качество работы лингвистов</u>	Мы отмечали <u>производительность</u>	"производительность" means output
The...experiments were concerned with conventional translation techniques	Опыты касались методов <u>обычных переводов</u>	Опыты были связаны с <u>общепринятой методикой перевода</u>	Not quite accurate translation
Teams of translators were used: one doing French-to-English and the other English-to-French.	Мы пользовались <u>ус- лугами /двух/ групп переводчиков: одной для переводов с французского на англ., а другой с англ. на франц.</u>	Были <u>использованы группы переводчиков: одни с франц. на англ., а другие с англ. на французский.</u>	пользоваться and использовать are not used with the object denoting a person
One experiment was intended to measure...translation	Один опыт был посвящен <u>определению качества...перевода</u>	Один опыт был посвящен <u>измерению перевода</u>	Inappropriate word (see page 2)
the speeches were delivered in about one hour	речи <u>продолжались приблизительно один час</u>	для <u>произнесения речи требовался приблизительно один час</u>	Inexact translation

Page 2

ENGLISH ORIGINAL	CORRECT TRANSLATION	WAS TRANSLATED	NOTE
our translation teams spent 37.6 hours	наши группы переводчиков <u>израсходовали</u> 37.6 часов	наши группы <u>переводчиков</u> <u>расходовали</u> 37.6 часов	Wrong verbal aspect!
the full translation rate...	<u>Скорость</u> <u>полного</u> <u>перевода</u>	<u>Полная</u> <u>скорость</u> <u>перевода</u>	Wrong translation
seven-eights of all passages...	в семи из <u>каждых</u> <u>восьми</u> отрывков	<u>среди</u> <u>семи</u> <u>из</u> <u>каждых</u> <u>восьми</u> отрывков	Inappropriate preposition
we reversed the target languages of the translation	мы <u>переменяли</u> <u>в</u> <u>группах</u> <u>переводчиков</u> <u>направление</u> <u>перевода</u>	мы <u>поставили</u> <u>перед</u> <u>группами</u> <u>переводчиков</u> <u>лингвистические</u> <u>задачи</u> <u>в</u> <u>обратном</u> <u>порядке</u>	The translation is obscure
to work in these unaccustomed directions	работать в этих <u>необычных</u> <u>направлениях</u>	работать в <u>необычных</u> <u>условиях</u>	Inaccurate translation
particularly for the English-speaking team	особенно в группе, <u>родным</u> <u>языком</u> <u>которой</u> <u>был</u> <u>английский</u>	особенно у тех, <u>которых</u> <u>в</u> <u>основном</u> <u>говорили</u> <u>по-английски</u>	Not clear!

ENGLISH ORIGINAL	CORRECT TRANSLATION	WAS TRANSLATED	NOTE
the "reversed target language" condition	<u>условие перевода в обратном направлении</u>	<u>условие "обратной" лингвистической задачи</u>	The expression is obscure
tended to slow down the translators	вело к замедлению работы переводчиков	имело тенденцию замедлить работу переводчиков	"иметь тенденцию" is usually used in connection with persons
A final brief experiment was done	В конце был также произведен короткий опыт	Был также произведен окончательный, но короткий опыт	Inappropriate word
re-translate back into the original languages	перевести обратно на язык оригинала	перевести обратно на первоначальный язык	Inexact translation
Similarly, retranslation back into English was slower than English to French translations	Подобным же образом, обратный перевод на английский... вод на английский...	Аналогичный обратный перевод на английский...	Inaccurate translation
for teleconferences requiring translation	для телесовещаний, нуждающихся в переводах...	для телесовещаний.....	The underlined words were not translated

ENGLISH ORIGINAL	CORRECT TRANSLATION	WAS TRANSLATED	NOTE
Because conventional full translations seemed to be unacceptable slow...	Так как обычный способ письменного перевода казался <u>недопустимо медленным</u> ,	Т.к. обычный способ письменного перевода был <u>недопустимо медлен</u> ,	"был" is not an equivalent of "seemed"
the medium of communication between participants was tele-type	телетайп служил средством связи <u>между участниками</u>	телетайп служил средством связи	Two underlined words were not translated
we used professional interpreters	мы пользовались <u>услугами устных переводчиков-профессионалов</u>	мы пользовались <u>устными переводчиками-профессионалами</u>	"пользоваться" is not used with the object denoting a person
We reversed the target languages of the sight translators	мы переменяли для зрительных переводчиков языки, <u>на которые они должны были переводить</u>	мы поставили перед зрительными переводчиками <u>обратную лингвистическую задачу</u>	The translation is obscure
The interpreter-subjects themselves predicted that they would perform poorly	Сами участники-переводчики <u>предсказывали, что они будут работать плохо</u>	Участвовавшие переводчики <u>сказали заранее, что у них будет плохая производительность</u>	"производительность" is not the same as "performance"

ENGLISH ORIGINAL	CORRECT TRANSLATION	WAS TRANSLATED	NOTE
Finally,	<u>В заключение,</u>	<u>Напоследок,</u>	A vulgate term
to re-translate material back into its original language	для обратного перевода материала на язык оригинала	для обратного перевода материала на его первоначальный язык	Not quite accurate translation
not all words were re-translated into their original identical form.	не все обратные введенные слова были идентичны со словами оригинала	не все обратные введенные слова были выражены идентично с их первоначальной формой	Not quite clear
Some interesting English examples of these changes are:	Вот несколько интересных английских примеров таких перемен	Среди интересных английских примеров можно упомянуть	Inexact translation
TELECOMMUNICATIONS MEDIA EXPERIMENTS	ЭКСПЕРИМЕНТЫ СО СРЕДСТВАМИ ТЕЛЕСВЯЗИ	ИСПЫТАНИЕ СРЕДСТВ ТЕЛЕСВЯЗИ	"испытание" (examination) does not correspond to the English "experiment." The correct Russian term is "телевидение" "телевизия"
or, in one case, television.	а в одном случае, телевидения	и в одном только случае - телевизи	
With the exception of television,	За исключением опыта с телевидением	За исключением опыта с телевизи	

ENGLISH ORIGINAL	CORRECT TRANSLATION	WAS TRANSLATED	NOTE
Participants played a game in which they had to allocate resources to a common group goal, as in an alliance, but with individual differences in the secret instructions given to each subject.	Участники вели игру, в которой они должны были распределять средства для общей групповой цели, как союзники, но с индивидуальными различиями в секретных инструкциях, данных каждому субъекту.	участники проводили игру, во время которой они должны были распределять ресурсы среди группы, объединенной общей целью, как например союзом, но каждому объекту были даны различные секретные инструкции.	Obscure translation Mistakes in spelling: ресурсы and in punctuation: например should be placed between 2 commas
Each conference was conducted with some urgency;	Каждая конференция проводилась с характером срочности.	Каждая конференция проводилась в рамках некоторой безотлагательности.	Awkward expression
That is, the group goal was to be met as soon as possible	т.е. групповая цель должна была быть выполнена как можно скорее.	т.е. требовалось удовлетворить групповую цель в возможно кратчайший срок.	One cannot say in Russian: удовлетворить цель
In the fourteen teleconferences...no two were identical...	Среди четырнадцати телеконференций...не было двух идентичных...	Среди четырнадцати телеконференций...не было ни одной идентичной...	Inaccurate translation

ENGLISH ORIGINAL	CORRECT TRANSLATION	WAS TRANSLATED	NOTE
Some conferences had a "party-line" or open network:	На некоторых конференциях были сети "с общим телефонным <u>про- водом</u> ", или открытые:	На некоторых конференциях были сети "с партийной <u>ли- нией</u> " или же <u>откры- тые</u> :	Wrong translation
Some networks were highly centralized and permitted a conferee to address himself...	Некоторые сети были сильно централизова-ны и позволяли <u>участ- нику совещания обра- щаться...</u>	Некоторые сети были сильно централизи-рованы, <u>гда участ- ник совещания мог обращаться...</u>	Poor construction The correct form is centralized
so that an inter- preter had to be used	так что пришлось <u>пользоваться услу- гой переводчика</u>	так что пришлось <u>повозоваться пере- водчиком</u>	See page 5
We also ran a few face-to-face prac- tice conferences	Мы провели также не-сколько тренировочных конференций, на кото-рых <u>присутствовали все участники /пере- говаривающихся групп/</u>	С целью тренировки мы имели также не-сколько конференций, на которых <u>все участ- ники присутствовали лицом к лицу.</u>	Not good expres- sion

ENGLISH ORIGINAL	CORRECT TRANSLATION	WAS TRANSLATED	NOTE
The need to compose messages...imposed a discipline on the conferees not necessary in direct voice communication	Необходимость составлять сообщения...налагала на участников конференций известную дисциплину, <u>не нужную при прямой устной связи</u>	Необходимость составлять сообщения...налагала на участников конференций известную дисциплину, что не всегда <u>относится к прямой устной связи</u>	Obscure translation
The...more terse language used in the teletype conferences	более сжатый язык, <u>употребляемый на теле-тайпных конференциях</u>	более сжатый язык, применяемый на теле-тайпных конференциях	Inappropriate verb
These meetings were much slower in reaching solutions	на этих совещаниях решения достигались гораздо медленнее	на этих совещаниях соглашения достигались гораздо медленнее	Inappropriate noun
some teletype networks provided true, two-way simultaneous transmission	некоторые теле-тайпные сети обещали <u>настоящую двухстороннюю одновременную передачу</u>	на некоторых теле-тайпных сетях существовала <u>настоящая возможность двухсторонней одновременной</u> передачи	Inaccurate translation

ENGLISH ORIGINAL	CORRECT TRANSLATION	WAS TRANSLATED	NOTE
...B was sending a message to <u>A</u>	В <u>высылал</u> свое <u>сообщение</u> по адресу <u>A</u>	В <u>высылал</u> свое <u>извещение</u> по адресу <u>A</u>	Inappropriate word
Fourth, the presence of a chairman in a centralized teletype network demonstrated the tremendous potential power of such an individual.	В-четвертых, <u>наличие</u> <u>председателя</u> в <u>централизованной</u> <u>теле-тайпной</u> <u>сети</u> <u>указывало</u> на <u>громдную</u> <u>потенциальную</u> <u>власть</u> <u>такого</u> <u>лица</u> .	В-четвертых, в <u>случае</u> <u>централизованной</u> <u>теле-тайпной</u> <u>сети</u> <u>лучалась</u> <u>картина</u> , <u>указывавшая</u> на <u>громдную</u> <u>потенциальную</u> <u>власть</u> ее <u>председателя</u> .	Not good translation The correct form is централизованной
The chairmen could also edit, delete, or change the sequence of messages going through their stations.	<u>Председатели</u> <u>также</u> <u>могли</u> <u>редактировать</u> , <u>вычеркивать</u> или <u>менять</u> <u>последовательность</u> <u>сообщений</u> , <u>идущих</u> <u>через</u> их <u>станции</u> .	он...на своей <u>станции</u> <u>мог</u> <u>редактировать</u> <u>проходившие</u> <u>сообщения</u> , их <u>задерживать</u> или же <u>менять</u> <u>очередность</u> .	Inaccurate translation
to insure that the addressee of a message has actually received...	чтобы <u>удостоверить</u> <u>ся</u> в <u>том</u> , что <u>адресат</u> <u>действительно</u> <u>получил</u> ...	для <u>проверки</u> <u>того</u> , что <u>адресат</u> <u>действительно</u> <u>получил</u> ...	Inexact translation

ENGLISH ORIGINAL	CORRECT TRANSLATION	WAS TRANSLATED	NOTE
acknowledgements... are easily trans- mitted by brief answers	подтверждения легко передать <u>в виде</u> коротких <u>ответов</u>	подтверждения пере- даются <u>в виде</u> корот- ких <u>ответов</u>	Mistake in spelling
face-to-face meetings also include non- verbal acknowledg- ments such as facial expressions or ges- tures.	заседания лицом к лицу включают под- тверждения <u>без слов</u> , т.е. <u>в виде</u> мимики лица или жестов.	при заседаниях лицом к лицу имеются также <u>бессловесные под-</u> <u>тверждения в виде</u> мимики или жестов.	The word "бессловесный" has the meaning "meek" or "dumb" and is used in poetry. Mistake in spel- ling of "в виде"
which put a non- English-speaking conferee at no ob- vious disadvantage from other conferees	что не ставило не говорящего по ан- глийски участника кон- ференции <u>в явно не-</u> <u>выгодное положение</u> по сравнению с другими	что исправляло не- выгодное <u>положение</u> не говорящего по- английски участни- ка конференции по сравнению с другими	Not clear!
experience with television	опыт с применением <u>телевидения</u>	опыт с применением <u>телевизи</u>	The correct form is "телевидения"

ENGLISH ORIGINAL	CORRECT TRANSLATION	WAS TRANSLATED	NOTE
remote negotiations...	лица, ведущие переговоры и находящиеся вдали,	лица... ведущие переговоры и находящиеся вдали,	Mistake in punctuation: a comma is omitted
or other non-verbal behavior	и другие проявления /чувств/ без слов...	и другие бессловесные проявления...	See above!
a good deal needs to be learned about how...	нужно будет еще много учиться, как...	придется узнать много, касающееся, того, как...	Inaccurate translation
experiments on two areas about which little is known	опытах в двух областях, еще мало изученных...	опытах, относящихся к двум мало изученным факторам...	Inexact translation
we believe we have demonstrated the feasibility	мы думаем, что мы наглядно показали осуществимость...	мы думаем, что нам удалось наглядно подтвердить осуществимость..	Inappropriate verb
a meeting...will be conducted as a teleconference with each of us in his own country at his own desk.	собрание...будет проводиться в виде телеконференции, где каждый из нас будет сидеть за своим письменным столом в своей собственной стране.	собрание...будет проводиться в виде телеконференции и каждый из нас будет сидеть за своим письменным столом в своей собственной стране.	Mistake in spelling (see above) Mistake in punctuation: a comma should be placed before "и"

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D		
(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)		
1. ORIGINATING ACTIVITY (Corporate author)		2a. REPORT SECURITY CLASSIFICATION
LATSEC, Inc. 7852 Ivanhoe Avenue La Jolla, CA 92037		Unclassified
3. REPORT TITLE		2b. GROUP
OPTIMIZATION OF SYSTRAN SYSTEM		N/A
4. DESCRIPTIVE NOTES (Type of report and inclusive dates)		
Final Technical Report 5 October 1970 - 5 October 1971		
5. AUTHOR(S) (First name, middle initial, last name)		
Dr. Peter P. Toma Dr. Ludek A. Kozlik Mr. Donald G. Perwin		
6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
April 1972	193	7
8a. CONTRACT OR GRANT NO.		9a. ORIGINATOR'S REPORT NUMBER(S)
F30602-71-C-0091 Job Order No. 30080000		
c. 04 001	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	
d. 31025F	RADC-TR-72-73	
10. DISTRIBUTION STATEMENT		
Approved for public release; distribution unlimited.		
11. SUPPLEMENTARY NOTES		12. SPONSORING MILITARY ACTIVITY
Report on Optimization of Russian-English Machine Translation System		Rome Air Development Center (IRDT) Griffiss Air Force Base, New York 13440
13. ABSTRACT		
This report describes a 12-month effort aimed at improvement of the SYSTRAN System for machine translation of Russian scientific and technical literature into English. In addition to programs that serve for automatic detection of errors in stored dictionaries and during the dictionary update process, routines have been written for improved treatment of homographs, polysemantic entries, adnominal genitives, numerals, proper names, abbreviations, tense sequences, adverbs and article insertion. A detailed account of results is given in the main body of the report. Along with an introductory description of the SYSTRAN System and some programming specifics, output translation samples are provided.		

DD FORM 1473

1 NOV 65

UNCLASSIFIED

Security Classification

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Computer Programming Machine Translation System Russian-English Translation						

UNCLASSIFIED

Security Classification